GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2001, 19:12:46 ; Search time 13.99 Seconds
(without alignments)
532.422 Million cell updates/sec

Title: US-09-170-042A-2
Perfect score: 1756
Sequence: 1 MENPSPAAALGKALCALLLA.....NGSPCPELEEEAECVPDNCV 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1756	100.0	331	2	US-08-799-173A-2	Sequence 2, Appli
2	1548.5	88.2	330	4	US-09-371-696-2	Sequence 2, Appli
3	464.5	26.5	802	1	US-07-862-021B-12	Sequence 12, Appl
4	464.5	26.5	802	1	US-08-313-288B-12	Sequence 12, Appl
5	464.5	26.5	802	5	PCT-US93-03164-12	Sequence 12, Appl
6	462.5	26.3	392	2	US-08-799-173A-7	Sequence 7, Appli
7	462.5	26.3	807	1	US-07-862-021B-10	Sequence 10, Appl
8	462.5	26.3	807	1	US-08-313-288B-10	Sequence 10, Appl
9	462.5	26.3	807	5	PCT-US93-03164-10	Sequence 10, Appl
10	420.5	23.9	132	4	US-09-022-238-2	Sequence 2, Appli
11	331	18.8	568	1	US-07-862-021B-14	Sequence 14, Appl
12	331	18.8	568	5	PCT-US93-03164-14	Sequence 14, Appl
13	309	17.6	53	2	US-08-799-173A-18	Sequence 18, Appl
14	189	10.8	37	4	US-09-022-238-3	Sequence 3, Appli
15	181	10.3	37	4	US-09-371-696-3	Sequence 3, Appli
16	130.5	7.4	56	1	US-07-862-021B-19	Sequence 19, Appl
17	130.5	7.4	56	5	PCT-US93-03164-19	Sequence 19, Appl
18	128.5	7.3	50	2	US-08-799-173A-14	Sequence 14, Appl
19	121.5	6.9	52	2	US-08-799-173A-12	Sequence 12, Appl
20	103	5.9	1251	5	PCT-US95-02251-3	Sequence 3, Appli
21	103	5.9	1252	1	US-08-199-780-3	Sequence 3, Appli
22	103	5.9	1252	2	US-08-316-650-3	Sequence 3, Appli
23	103	5.9	1253	3	US-08-479-722B-4	Sequence 4, Appli
24	99.5	5.7	812	1	US-08-446-794A-4	Sequence 4, Appli
25	96.5	5.5	812	1	US-08-446-794A-2	Sequence 2, Appli
26	96.5	5.5	812	1	US-08-750-007-3	Sequence 3, Appli
27	96.5	5.5	812	2	US-08-945-024-2	Sequence 2, Appli
28	92.5	5.3	912	5	PCT-US95-03747-2	Sequence 2, Appli
29	91.5	5.2	52	2	US-08-799-173A-10	Sequence 10, Appl
30	91.5	5.2	56	1	US-07-862-021B-17	Sequence 17, Appl
31	91.5	5.2	56	5	PCT-US93-03164-17	Sequence 17, Appl
32	90	5.1	53	2	US-08-799-173A-9	Sequence 9, Appli
33	89.5	5.1	56	1	US-07-862-021B-16	Sequence 16, Appl
34	89.5	5.1	56	5	PCT-US93-03164-16	Sequence 16, Appl
35	89.5	5.1	1156	3	US-08-996-083-1	Sequence 1, Appli
36	89.5	5.1	1156	4	US-09-429-516-1	Sequence 1, Appli
37	89.5	5.1	1156	4	US-09-429-516-3	Sequence 3, Appli
38	89	5.1	1257	4	US-09-220-641-3	Sequence 3, Appli
39	87.5	5.0	1184	2	US-08-918-914-1	Sequence 1, Appli
40	87.5	5.0	1184	3	US-08-996-083-3	Sequence 3, Appli
41	86.5	4.9	224	4	US-08-871-572B-11	Sequence 11, Appl
42	85.5	4.9	692	4	US-09-003-574-31	Sequence 31, Appl
43	85.5	4.9	692	4	US-09-003-570-31	Sequence 31, Appl
44	85.5	4.9	695	2	US-08-701-240-4	Sequence 4, Appli
45	85.5	4.9	695	4	US-09-138-236-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT 1
US-08-799-173A-2
; Sequence 2, Application US/08799173A
; Patent No. 5871969
; GENERAL INFORMATION:
;   APPLICANT: HASTINGS, GREGG,
;   APPLICANT: PATRICK J. DILLON
;   TITLE OF INVENTION: HUMAN NEURONAL ATTACHMENT FACTOR-1
;   NUMBER OF SEQUENCES: 18
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: HUMAN GENOME SCIENCES,INC.
;     STREET: 9410 KEY WEST AVENUE
;     CITY: ROCKVILLE
;     STATE: MD
;     COUNTRY: USA
;     ZIP: 20850
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/799,173A
;     FILING DATE: 11-FEB-1997
;     CLASSIFICATION: 536
;   ATTORNEY/AGENT INFORMATION:
;     NAME: BROOKES, ANDERS A.
;     REGISTRATION NUMBER: 36,373
;     REFERENCE/DOCKET NUMBER: PF226
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (301) 309-8504
;     TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 331 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-799-173A-2


  Query Match           100.0%;  Score 1756;  DB 2;  Length 331;
  Best Local Similarity 100.0%;  Pred. No. 3e-166;
  Matches  331;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQY 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQY 60

Qy      61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120

Qy     121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180

Qy     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240

Qy     241 LVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 LVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300

Qy     301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
                |||||||||||||||||||||||||||||
Db     301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331


; RESULT   2
(  09-371-696-2
; Sequence 2, Application US/09371696
; Patent No. 6287777
; GENERAL INFORMATION:
;  APPLICANT: Sytkowski, Arthur J.
;  APPLICANT: Yang, Meiheng
;  TITLE OF INVENTION: NOVEL NPG-1 GENE THAT IS DIFFERENTIALLY EXPRESSED IN PROSTATE
;  TITLE OF INVENTION: TUMORS
;  FILE REFERENCE: 01948/053002
;  CURRENT APPLICATION NUMBER: US/09/371,696
;  CURRENT FILING DATE: 1999-08-10
;  EARLIER APPLICATION NUMBER: US 09/022,238
;  EARLIER FILING DATE: 1998-02-11
;  EARLIER APPLICATION NUMBER: US 08/644,326
;  EARLIER FILING DATE: 1996-05-10
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 330
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-371-696-2


  Query Match           88.2%;  Score 1548.5;  DB 4;  Length 330;
  Best Local Similarity  87.6%;  Pred. No. 1.2e-145;
  Matches  297;  Conservative    6;  Mismatches   19;  Indels   17;  Gaps    2;

        1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQY 60
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db      1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARAPAKYSITFTGKWSQTAFPKQY 60

Qy     61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120

Qy    121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180
          | |||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db    121 HEVFSAPAVPSGTGQTSAELEVQRKHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180

Qy    181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240
          |||||||||||||||||||||||||||||||||||||||||:|||||||||
Db    181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPASSFYYPRLKACSH----- 235

Qy    241 LVRLRQ--------SPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGH 292
              ||         |:   |  |    :|| :|||||||||||||||||||||||||||||
Db    236 ----RQGDTGAAATEPQGLHPSRPSPAQQDNALVDSASVPETPLDCEVSLWSSWGLCGGH 291

Qy    293 CGRLGTKSRTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
          |||||:|||| | ||||||||||||||||||||| ||||
Db    292 CGRLGSKSRTPYARVQPANNGSPCPELEEEAECVADNCV 330


RESULT   3
US-07-862-021B-12
; Sequence 12, Application US/07862021B
; Patent No. 5279966
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M
;    APPLICANT:  Klar, Avihu
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/862,021B
;      FILING DATE:  19920405
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  802 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-862-021B-12


  Query Match           26.5%;  Score 464.5;  DB 1;  Length 802;
  Best Local Similarity  33.2%;  Pred. No. 1.6e-37;
  Matches  111;  Conservative   52;  Mismatches  144;  Indels   27;  Gaps   10;

Qy      9 ALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQ 68
          :| | :|    |: |   :|      | |   ||| :|| | ||:   || :|  |
Db    171 SLTKRICEQDSASEGVTDKP---TLDCCACGTAKYRLTFYGNWSEKTHPKDFP--RRTNH 225

Qy     69 WSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEI-EAAGEALQSVHAVFSAP 127
          ||:::|::|| :| :|    | | |::  || |    : :|| : : | |  : |     |
Db    226 WSAIIGSSHSKNYILWEYGGYASEGVKQVAELGSPVKMEEEIRQQSDEVLTVIKAKAQWP 285

Qy    128 AVP--SGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGD-RWREQAALDL 184
          |    :     |||  | |   |:||:  : ||||| ||: : |||  |  | ::   ||
Db    286 AWQPLNVRAAPSAEFSVDRHRHLMSFLTMLGPSPDWNVGLSAEDLCTKDCGWVQKVVQDL 345

Qy    185 YPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVTLVRL 244
           |:|||||| |: ||| |:||: :  :||    || : || |   ::  :||| | |:
Db    346 IPWDAGTDSGVTYESPNKPTVPQEKIRPLTSL--DHPQSPFYDPEGGSIKLVARVVLERI 403

Qy    245 RQSPRA--FIPP------APVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRL 296
           :     |:|       | : |     |          :||  |  | || |  |
Db    404 ARKGEQCNFVPDNIDDIVADLAPEEKEE-------DDTPETCIYSNWSPWSACSSSTCEK 456

Qy    297 GTKSRTRYVRVQPANNGSPCPELEEEAECVPDNC 330
          | : | | :: |  :   |||: ::   |:   |
Db    457 GKRMRQRMLKAQ-LDLSVPCPDTQDFQPCMGPGC 489
```

```
RESULT   4
US-08-313-288B-12
; Sequence 12, Application US/08313288B
; Patent No. 5750502
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M. and Avihu Klar
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/313,288B
;      FILING DATE:  January 5, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028-A-PCT-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0526
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  802 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-313-288B-12


  Query Match             26.5%;  Score 464.5;  DB 1;  Length 802;
  Best Local Similarity   33.2%;  Pred. No. 1.6e-37;
  Matches  111;  Conservative   52;  Mismatches  144;  Indels   27;  Gaps   10;

Qy       9 ALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQ 68
           :| | :|    |: |   :|      | |   ||| :|| | ||:   || :|  |
Db     171 SLTKRICEQDSASEGVTDKP---TLDCCACGTAKYRLTFYGNWSEKTHPKDFP--RRTNH 225

Qy      69 WSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEI-EAAGEALQSVHAVFSAP 127
           ||:::|::|| :| :|    | | |::  || |    : :|| : : | |  : |    |
Db     226 WSAIIGSSHSKNYILWEYGGYASEGVKQVAELGSPVKMEEEIRQQSDEVLTVIKAKAQWP 285

Qy     128 AVP--SGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGD-RWREQAALDL 184
           |    :     |||  | |   |:||:  : ||||| ||: : |||  |  | ::   ||
Db     286 AWQPLNVRAAPSAEFSVDRHRHLMSFLTMLGPSPDWNVGLSAEDLCTKDCGWVQKVVQDL 345

Qy     185 YPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVTLVRL 244
           |:|||||| |: |||  |:||: :  :||    || : || |   ::  :||| | |:
Db     346 IPWDAGTDSGVTYESPNKPTVPQEKIRPLTSL--DHPQSPFYDPEGGSIKLVARVVLERI 403

Qy     245 RQSPRA--FIPP------APVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRL 296
           :     |:|       | : |    |         :||  |  | || |  |
Db     404 ARKGEQCNFVPDNIDDIVADLAPEEKEE-------DDTPETC1YSNWSPWSACSSSTCEK 456

Qy     297 GTKSRTRYVRVQPANNGSPCPELEEEAECVPDNC 330
           | : | | :: |  :   |||: ::   |:  |
Db     457 GKRMRQRMLKAQ-LDLSVPCPDTQDFQPCMGPGC 489


RESULT   5
PCT-US93-03164-12
```

```
; Sequence 12, Application PC/TUS9303164
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M
;    APPLICANT:  Klar, Avihu
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/03164
;      FILING DATE:  19930402
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  802 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US93-03164-12


  Query Match             26.5%;  Score 464.5;  DB 5;  Length 802;
  Best Local Similarity   33.2%;  Pred. No. 1.6e-37;
  Matches  111;  Conservative   52;  Mismatches  144;  Indels   27;  Gaps   10;

Qy       9 ALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQ 68
           :| | :|    |: |   :|      | |   ||| :|| | ||:   || :|  |
Db     171 SLTKRICEQDSASEGVTDKP---TLDCCACGTAKYRLTFYGNWSEKTHPKDFP--RRTNH 225

Qy      69 WSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEI-EAAGEALQSVHAVFSAP 127
           ||:::|::|| :| :|    | | |::  || |    : :|| : : | |  : |    |
Db     226 WSAIIGSSHSKNYILWEYGGYASEGVKQVAELGSPVKMEEEIRQQSDEVLTVIKAKAQWP 285

Qy     128 AVP--SGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGD-RWREQAALDL 184
           |    :     |||  | |   |:||:  : ||||| ||: : |||  |  | ::   ||
Db     286 AWQPLNVRAAPSAEFSVDRHRHLMSFLTMLGPSPDWNVGLSAEDLCTKDCGWVQKVVQDL 345

Qy     185 YPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVTLVRL 244
           |:|||||| |: |||  |:||: :  :||    || : || |   ::  :||| | |:
Db     346 IPWDAGTDSGVTYESPNKPTVPQEKIRPLTSL--DHPQSPFYDPEGGSIKLVARVVLERI 403

Qy     245 RQSPRA--FIPP------APVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRL 296
           :     |:|       | : |    |         :||  |  | || |  |
Db     404 ARKGEQCNFVPDNIDDIVADLAPEEKEE-------DDTPETCIYSNWSPWSACSSSTCEK 456

Qy     297 GTKSRTRYVRVQPANNGSPCPELEEEAECVPDNC 330
           | : | | :: |  :   |||: ::   |:  |
Db     457 GKRMRQRMLKAQ-LDLSVPCPDTQDFQPCMGPGC 489


RESULT   6
```

```
US-08-799-173A-7
; Sequence 7, Application US/08799173A
; Patent No. 5871969
;  GENERAL INFORMATION:
;    APPLICANT:  HASTINGS, GREGG,
;    APPLICANT:  PATRICK J. DILLON
;    TITLE OF INVENTION:  HUMAN NEURONAL ATTACHMENT FACTOR-1
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HUMAN GENOME SCIENCES,INC.
;      STREET:  9410 KEY WEST AVENUE
;      CITY:  ROCKVILLE
;      STATE:  MD
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/799,173A
;      FILING DATE:  11-FEB-1997
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROOKES, ANDERS A.
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PF226
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  392 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MQLECULE TYPE:  protein
US-08-799-173A-7

  Query Match             26.3%;  Score 462.5;  DB 2;  Length 392;
  Best Local Similarity   34.2%;  Pred. No. 9e-38;
  Matches  116;  Conservative   49;  Mismatches  137;  Indels   37;  Gaps   12;

Qy       9 ALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQ 68
Db      26 SLTKKLCEQDPTLDGVTDRPI---LDCCACGTAKYRLTFYGNWSEKTHPKDYP--RRANH 80

Qy      69 WSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEI-EAAGEALQSVHAVFSAP 127
Db      81 WSAIIGGSHSKNYVLWEYGGYASEGVKQVAELGSPVKMEEEIRQQSDEVLTVIKAKAQWP 140

Qy     128 --------AVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGD-RWRE 178
Db     141 SWQPVNVRAAP------SAEFSVDRTRHLMSFLTMMGPSPDWNVGLSAEDLCTKECGWVQ 194

Qy     179 QAALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIAR 238
Db     195 KVVQDLIPWDAGTDSGVTYESPNKPTIPQEKIRPLTSL--DHPQSPFYDPEGGSITQVAR 252

Qy     239 VTLVRL-RQSPRAFIPPAPVLPSRDNEIVDSASVPE------TPLDCEVSLWSSWGLCGG 291
Db     253 VVIERIARKGEQCNIVPDNV----DDIVADLA--PEEKDEDDTPETCIYSNWSPWSACSS 306

Qy     292 HCGRLGTKSRTRYVRVQPANNGSPCPELEEEAECVPDNC 330
Db     307 STCEKGKRMRQRMLKAQ-LDLSVPCPDTQDFQPCMGPGC 344

RESULT   7
US-07-862-021B-10
; Sequence 10, Application US/07862021B
; Patent No. 5279966
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M
;    APPLICANT:  Klar, Avihu
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/862,021B
;      FILING DATE:  19920405
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  807 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-862-021B-10

  Query Match             26.3%;  Score 462.5;  DB 1;  Length 807;
  Best Local Similarity   34.2%;  Pred. No. 2.6e-37;
  Matches  116;  Conservative   49;  Mismatches  137;  Indels   37;  Gaps   12;

Qy       9 ALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQ 68
Db     176 SLTKKLCEQDPTLDGVTDRPI---LDCCACGTAKYRLTFYGNWSEKTHPKDYP--RRANH 230

Qy      69 WSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEI-EAAGEALQSVHAVFSAP 127
Db     231 WSAIIGGSHSKNYVLWEYGGYASEGVKQVAELGSPVKMEEEIRQQSDEVLTVIKAKAQWP 290

Qy     128 --------AVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGD-RWRE 178
Db     291 SWQPVNVRAAP------SAEFSVDRTRHLMSFLTMMGPSPDWNVGLSAEDLCTKECGWVQ 344

Qy     179 QAALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIAR 238
Db     345 KVVQDLIPWDAGTDSGVTYESPNKPTIPQEKIRPLTSL--DHPQSPFYDPEGGSITQVAR 402

Qy     239 VTLVRL-RQSPRAFIPPAPVLPSRDNEIVDSASVPE------TPLDCEVSLWSSWGLCGG 291
Db     403 VVIERIARKGEQCNIVPDNV----DDIVADLA--PEEKDEDDTPETCIYSNWSPWSACSS 456

Qy     292 HCGRLGTKSRTRYVRVQPANNGSPCPELEEEAECVPDNC 330
Db     457 STCEKGKRMRQRMLKAQ-LDLSVPCPDTQDFQPCMGPGC 494

RESULT   8
US-08-313-288B-10
; Sequence 10, Application US/08313288B
; Patent No. 5750502
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M. and Avihu Klar
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/313,288B
;      FILING DATE:  January 5, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028-A-PCT-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0526
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  807 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-313-288B-10

  Query Match             26.3%;  Score 462.5;  DB 1;  Length 807;
  Best Local Similarity   34.2%;  Pred. No. 2.6e-37;
  Matches  116;  Conservative   49;  Mismatches  137;  Indels   37;  Gaps   12;

Qy       9 ALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQ 68
           :| | ||      |   :|:     | |   ||| :|| | ||:   || ||  |
Db     176 SLTKKLCEQDPTLDGVTDRPI---LDCCACGTAKYRLTFYGNWSEKTHPKDYP--RRANH 230

Qy      69 WSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEI-EAAGEALQSVHAVFSAP 127
           ||:::| :|| :| :|    | | |::  || |    : :|| : : | |  : |    |
Db     231 WSAIIGGSHSKNYVLWEYGGYASEGVKQVAELGSPVKMEEEIRQQSDEVLTVIKAKAQWP 290

Qy     128 --------AVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGD-RWRE 178
                 | |      |||  | |   |:||:  : ||||| ||: : |||  :  | :
Db     291 SWQPVNVRAAP------SAEFSVDRTRHLMSFLTMMGPSPDWNVGLSAEDLCTKECGWVQ 344

Qy     179 QAALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIAR 238
           :   || |:|||||| |: |||  ||||: :  :||    || : || |   ::  :||
Db     345 KVVQDLIPWDAGTDSGVTYESPNKPTIPQEKIRPLTSL--DHPQSPFYDPEGGSITQVAR 402

Qy     239 VTLVRL-RQSPRAFIPPAPVLPSRDNEIVDSASVPE------TPLDCEVSLWSSWGLCGG 291
           | : |: |:  :  | |  |    |: : | |  ||      ||  |  | || |  |
Db     403 VVIERIARKGEQCNIVPDNV----DDIVADLA--PEEKDEDDTPETCIYSNWSPWSACSS 456

Qy     292 HCGRLGTKSRTRYVRVQPANNGSPCPELEEEAECVPDNC 330
             | : | | :: |  :   |||: ::   |:   |
Db     457 STCEKGKRMRQRMLKAQ-LDLSVPCPDTQDFQPCMGPGC 494

RESULT  9
PCT-US93-03164-10
; Sequence 10, Application PC/TUS9303164
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M
;    APPLICANT:  Klar, Avihu
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/03164
;      FILING DATE:  19930402
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  807 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US93-03164-10

  Query Match             26.3%;  Score 462.5;  DB 5;  Length 807;
  Best Local Similarity   34.2%;  Pred. No. 2.6e-37;
  Matches  116;  Conservative   49;  Mismatches  137;  Indels   37;  Gaps   12;

Qy       9 ALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQ 68
           :| | ||      |   :|:     | |   ||| :|| | ||:   || ||  |
Db     176 SLTKKLCEQDPTLDGVTDRPI---LDCCACGTAKYRLTFYGNWSEKTHPKDYP--RRANH 230

Qy      69 WSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEI-EAAGEALQSVHAVFSAP 127
           ||:::| :|| :| :|    | | |::  || |    : :|| : : | |  : |    |
Db     231 WSAIIGGSHSKNYVLWEYGGYASEGVKQVAELGSPVKMEEEIRQQSDEVLTVIKAKAQWP 290

Qy     128 --------AVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGD-RWRE 178
                 | |      |||  | |   |:||:  : ||||| ||: : |||  :  | :
Db     291 SWQPVNVRAAP------SAEFSVDRTRHLMSFLTMMGPSPDWNVGLSAEDLCTKECGWVQ 344

Qy     179 QAALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIAR 238
           :   || |:|||||| |: |||  ||||: :  :||    || : || |   ::  :||
Db     345 KVVQDLIPWDAGTDSGVTYESPNKPTIPQEKIRPLTSL--DHPQSPFYDPEGGSITQVAR 402

Qy     239 VTLVRL-RQSPRAFIPPAPVLPSRDNEIVDSASVPE------TPLDCEVSLWSSWGLCGG 291
           | : |: |:  :  | |  |    |: : | |  ||      ||  |  | || |  |
Db     403 VVIERIARKGEQCNIVPDNV----DDIVADLA--PEEKDEDDTPETCIYSNWSPWSACSS 456

Qy     292 HCGRLGTKSRTRYVRVQPANNGSPCPELEEEAECVPDNC 330
             | : | | :: |  :   |||: ::   |:   |
Db     457 STCEKGKRMRQRMLKAQ-LDLSVPCPDTQDFQPCMGPGC 494

RESULT  10
US-09-022-238-2
; Sequence 2, Application US/09022238
; Patent No. 6177244
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Sytkowski, Arthur J. and Yang, Meiheng
;    TITLE OF INVENTION:  A novel NPG-1 Gene that is differentially expressed in prosta
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/022,238
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/644,326
;      FILING DATE:  10-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maravic-Magovcevic, Ivana
;      REGISTRATION NUMBER:  P-43,338
;      REFERENCE/DOCKET NUMBER:  NER-262CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)742-4214
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  132 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-022-238-2


  Query Match             23.9%;  Score 420.5;  DB 4;  Length 132;
  Best Local Similarity   65.4%;  Pred. No. 2.8e-34;
  Matches   83;  Conservative    5;  Mismatches   22;  Indels   17;  Gaps    2;

Qy     200 PNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVTLVRLRQ--------SPRAF 251
           |     | ||||||||||||||:|||||||||           ||          |:
Db      15 PQLRHHPADTVTEITSSSPSHPASSFYYPRLKACSH---------RQGDTGAAATEPQGL 65

Qy     252 IPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKSRTRYVRVQPAN 311
           |  |    :|| :|||||||||||||||||||||||||||||||:|||| | ||||||
Db      66 HPSRPSPAQQDNALVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGSKSRTPYARVQPAN 125

Qy     312 NGSPCPE 318
           |||||||
Db     126 NGSPCPE 132


RESULT  11
US-07-862-021B-14
; Sequence 14, Application US/07862021B
; Patent No. 5279966
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M
;    APPLICANT:  Klar, Avihu
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/862,021B
;      FILING DATE:  19920405
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  568 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-862-021B-14


  Query Match             18.8%;  Score 331;  DB 1;  Length 568;
  Best Local Similarity   33.1%;  Pred. No. 1.7e-24;
  Matches   87;  Conservative   40;  Mismatches  116;  Indels   20;  Gaps    9;

Qy      78 SSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAG-EALQSVHAVFSAPAVP--SGTG 134
           | :| :|   |  |:|::  || |    : :||   | | |  : |    ||    :
Db       1 SGEYVLWSMRQ-ASDGVKQVAELGSPVKMEEEIRQKGDEVLTVIKAKAQWPAWQPLNVRA 59

Qy     135 QTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGD-RWREQAALDLYPYDAGTDS 193
             |||  | |   |:||:  : ||||| ||: | |||  :  | ::   || |:||||||
Db      60 APSAEFSVDRSRHLMSFLAMMGPSPDWNVGLTSEDLCTKECGWVQKVVQDLIPWDAGTDS 119

Qy     194 GFTFSSPNFATIPQDTVTEITS-SSPSHPANSFYYPRLKALPPIARVTLVRLRQSPRAFI 252
           | |: |||  ||||| :  :||   |  |: :    |   : |||||  : |: :
Db     120 GVTYESPNKPTIPQDKIRPLTSLDHPQSPSMT----RGGPIIPIARVVIERIARKGE--- 172

Qy     253 PPAPVLPSRDNEIV-----DSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKSRTRYVRV 307
             ::|   ::||      :     :||  |  | || | |       | : | | ::
Db     173 -QCNIIPDNVDDIVADLVTEEKDEDDTPETCIYSNWSPWSACSSATCDKGKRMRQRMLKA 231

Qy     308 QPANNGSPCPELEEEAECVPDNC 330
           |  :   |||: ::   |:   |
Db     232 Q-LDLSVPCPDTQDFEPCMGPGC 253


RESULT  12
PCT-US93-03164-14
; Sequence 14, Application PC/TUS9303164
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M
;    APPLICANT:  Klar, Avihu
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/03164
```

```
;      FILING DATE:  19930402
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  568 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;-US93-03164-14

  Query Match             18.8%;  Score 331;  DB 5;  Length 568;
  Best Local Similarity   33.1%;  Pred. No. 1.7e-24;
  Matches   87;  Conservative   40;  Mismatches  116;  Indels   20;  Gaps    9;

Qy      78 SSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAG-EALQSVHAVFSAPAVP--SGTG 134
           | :| :|   |  |:|::  || |    : :||   | | |  : |    ||    :
Db       1 SGEYVLWSMRQ-ASDGVKQVAELGSPVKMEEEIRQKGDEVLTVIKAKAQWPAWQPLNVRA 59

Qy     135 QTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGD-RWREQAALDLYPYDAGTDS 193
           .    |||  | |   |:||:  : ||||| ||: | |||  :  | ::   || |:||||||
Db      60 APSAEFSVDRSRHLMSFLAMMGPSPDWNVGLTSEDLCTKECGWVQKVVQDLIPWDAGTDS 119

Qy     194 GFTFSSPNFATIPQDTVTEITS-SSPSHPANSFYYPRLKALPPIARVTLVRLRQSPRAFI 252
           | |: |||  ||||| :  :||   |  |: :    |   : ||||| : |: :
Db     120 GVTYESPNKPTIPQDKIRPLTSLDHPQSPSMT----RGGPIIPIARVVIERIARKGE--- 172

Qy     253 PPAPVLPSRDNEIV-----DSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKSRTRYVRV 307
               ::|   ::||     :     :|| |  | || |  |       | : | | ::
Db     173 -QCNIIPDNVDDIVADLVTEEKDEDDTPETCIYSNWSPWSACSSATCDKGKRMRQRMLKA 231

Qy     308 QPANNGSPCPELEEEAECVPDNC 330
           |  :   |||: ::   |:   |
Db     232 Q-LDLSVPCPDTQDFEPCMGPGC 253


RESULT  13
US-08-799-173A-18
; Sequence 18, Application US/08799173A
; Patent No. 5871969
;  GENERAL INFORMATION:
;    APPLICANT:  HASTINGS, GREGG,
;    APPLICANT:  PATRICK J. DILLON
;    TITLE OF INVENTION:  HUMAN NEURONAL ATTACHMENT FACTOR-1
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HUMAN GENOME SCIENCES,INC.
;      STREET:  9410 KEY WEST AVENUE
;      CITY:  ROCKVILLE
;      STATE:  MD
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/799,173A
;      FILING DATE:  11-FEB-1997
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROOKES, ANDERS A.
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PF226
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-799-173A-18


  Query Match             17.6%;  Score 309;  DB 2;  Length 53;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-24;
  Matches   53;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     278 CEVSLWSSWGLCGGHCGRLGTKSRTRYVRVQPANNGSPCPELEEEAECVPDNC 330
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CEVSLWSSWGLCGGHCGRLGTKSRTRYVRVQPANNGSPCPELEEEAECVPDNC 53


RESULT  14
US-09-022-238-3
; Sequence 3, Application US/09022238
; Patent No. 6177244
;  GENERAL INFORMATION:
;    APPLICANT:  Sytkowski, Arthur J. and Yang, Meiheng
;    TITLE OF INVENTION:  A novel NPG-1 Gene that is differentially expressed in pros
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/022,238
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/644,326
;      FILING DATE:  10-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maravic-Magovcevic, Ivana
;      REGISTRATION NUMBER:  P-43,338
;      REFERENCE/DOCKET NUMBER:  NER-262CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)742-4214
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  37 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-09-022-238-3


  Query Match             10.8%;  Score 189;  DB 4;  Length 37;
  Best Local Similarity   91.4%;  Pred. No. 4.1e-12;
  Matches   32;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy     261 RDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGR 295
           :|| :|||||||||||||||||||||||||||||||
Db       3 QDNALVDSASVPETPLDCEVSLWSSWGLCGGHCGR 37


RESULT  15
US-09-371-696-3
; Sequence 3, Application US/09371696
; Patent No. 6287777
; GENERAL INFORMATION:
;  APPLICANT: Sytkowski, Arthur J.
;  APPLICANT: Yang, Meiheng
;  TITLE OF INVENTION: NOVEL NPG-1 GENE THAT IS DIFFERENTIALLY EXPRESSED IN PROSTATE
;  TITLE OF INVENTION: TUMORS
;  FILE REFERENCE: 01948/053002
;  CURRENT APPLICATION NUMBER: US/09/371,696
;  CURRENT FILING DATE: 1999-08-10
;  EARLIER APPLICATION NUMBER: US 09/022,238
;  EARLIER FILING DATE: 1998-02-11
;  EARLIER APPLICATION NUMBER: US 08/644,326
;  EARLIER FILING DATE: 1996-05-10
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 37
;   TYPE: PRT
;   ORGANISM: Homo sqpiens
US-09-371-696-3


  Query Match             10.3%;  Score 181;  DB 4;  Length 37;
  Best Local Similarity   88.6%;  Pred. No. 2.6e-11;
  Matches   31;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     261 RDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGR 295
           :|| :|||||||||||||||||||| ||||||||
Db       3 QDNALVDSASVPETPLDCEVSLWSSWELCGGHCGR 37


Search completed: December 10, 2001, 20:00:45
Job time: 2879 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2001, 13:38:51 ; Search time 75.12 Seconds
(without alignments)
326.388 Million cell updates/sec

Title: US-09-170-042A-2
Perfect score: 1756
Sequence: 1 MENPSPAAALGKALCALLLA.....NGSPCPELEEEAECVPDNCV 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1756	100.0	331	20	AAW92460	Human NAF-1 protei
2	1748	99.5	331	18	AAW23663	Human neuronal att
3	1747	99.5	331	20	AAW83328	Human mindin polyp
4	1746	99.4	331	22	AAM38872	Human polypeptide
5	1746	99.4	444	22	AAM40658	Human polypeptide
6	1744	99.3	331	21	AAY79561	Cancer specific pr
7	1742	99.2	331	22	AAB82472	Human extracellula
8	1741	99.1	331	19	AAW70589	Adhesion-modulatin
9	1741	99.1	331	20	AAY41721	Human PRO866 prote
10	1741	99.1	331	21	AAB44277	Human PRO866 (UNQ4
11	1741	99.1	331	21	AAB33465	Human PRO866 prote
12	1741	99.1	331	21	AAY95349	Human PRO866 antit
13	1740	99.1	385	22	AAG75576	Human colon cancer
14	1329	75.7	249	21	AAB34753	Human secreted pro
15	1101.5	62.7	299	21	AAY73490	Human secreted pro
16	583	33.2	290	20	AAW83329	Human mindin-relat
17	464.5	26.5	802	14	AAR44242	F-spondin of chick
18	462.5	26.3	392	20	AAW92461	Rat F-spondin prot
19	462.5	26.3	807	14	AAR44241	F-spondin (FP5-9).
20	462.5	26.3	807	20	AAY04264	Rat vascular smoot
21	462.5	26.3	807	21	AAB13074	Vascular endotheli
22	460.5	26.2	807	20	AAY04262	Human vascular smo
23	460.5	26.2	807	20	AAY04263	Bovine vascular sm
24	460.5	26.2	807	21	AAB13072	Vascular endotheli
25	460.5	26.2	807	21	AAB13073	Vascular endotheli
26	460.5	26.2	807	22	AAB74401	Human ovarian canc
27	420.5	23.9	132	20	AAY34113	Human NPG-1 protei
28	351	20.0	63	21	AAB27878	Protein fragment e
29	331	18.8	568	14	AAR44243	Xenopus F-spondin.
30	309	17.6	53	20	AAW92469	Human NAF-1 protei
31	277	15.8	105	21	AAB34693	Human secreted pro
32	250	14.2	63	21	AAB27877	Sequence homologou
33	155.5	8.9	182	21	AAB58354	Lung cancer associ
34	155.5	8.9	343	22	AAB93099	Human protein sequ
35	134	7.6	1551	22	AAM41081	Human polypeptide
36	134	7.6	1588	22	AAM39295	Human polypeptide
37	134	7.6	1588	22	AAB20155	Secreted protein S
38	128.5	7.3	50	20	AAW92468	Rat F-spondin FSP-
39	126.5	7.2	264	22	AAB48109	Mouse TANGO 204 po
40	126.5	7.2	264	22	AAB48130	Mouse TANGO 204 va
41	124.5	7.1	264	22	AAB48131	Mouse TANGO 204 va
42	123.5	7.0	264	22	AAB48132	Mouse TANGO 204 va
43	121.5	6.9	52	20	AAW92466	Rat F-spondin FSP-
44	116.5	6.6	242	22	AAM42006	Human polypeptide
45	116.5	6.6	264	22	AAB48105	Human TANGO 204 po

ALIGNMENTS

RESULT 1
AAW92460
ID AAW92460 standard; Protein; 331 AA.
XX
AC AAW92460;
XX
DT 21-APR-1999 (first entry)
XX
DE Human NAF-1 protein.
XX
KW NAF-1; neuronal attachment factor-1; F-spondin analogue; treatment;
KW spinal cord injury; peripheral nerves damage; neural cell adhesion;
KW neurite extension; tumour cell metastasis; inhibitor; mobility; disease;
KW endothelial cell proliferation; tumour neovascularisation; haemostasis;
KW angiostatic agent; wound healing; diagnostic; neurotrophic; anticancer;
KW antimetastatic; anti-angiogenic; antimalarial; human.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..23
FT /label= signal_peptide
FT Protein 24..331
XX
PN US5871969-A.
XX
PD 16-FEB-1999.
XX
PF 12-FEB-1997; 97US-0799173.
XX
PR 12-FEB-1996; 96US-0011519.
PR 12-FEB-1997; 97US-0799173.
XX
PA (HUMA-) HUMAN GENOME SCI INC.

```
XX
PI   Dillon PJ,  Hastings G;
XX
DR   WPI; 1999-166644/14.
DR   N-PSDB; AAX02019.
XX
PT   Nucleic acid encoding human neuronal attachment factor-1  - used to
PT   e.g. treat spinal cord injuries, and inhibit tumor cell metastasis
PT   and neovascularization
XX
PS   Claim 1a; Fig 1; 29pp; English.
XX
CC   This sequence represents a novel human neuronal attachment factor-1,
CC   NAF-1. NAF-1 is an analogue of rat F-spondin and is used to treat spinal
CC   cord injuries and damage to peripheral nerves (by promoting neural cell
CC   adhesion and neurite extension), to inhibit tumor cell metastasis
CC   (particularly in small cell and breast carcinoma) and endothelial cell
CC   proliferation, adhesion and mobility, to reduce tumor neovascularisation,
CC   as angiostatic agents for tumor cells, to promote wound healing and to
CC   modulate hemostasis. NAF-1 may also be used to identify treatments and
CC   diagnoses for human disease. Fragments of the nucleic acid that do not
CC   encode NAF-1 peptides are useful as probes to isolate the NAF-1 gene, its
CC   allelic variants, full-length cDNA or related sequences, in chromosomal
CC   location by in situ hybridization or in Northern blotting, and as
CC   diagnostic probes or primers. The protein has neurotrophic, anticancer,
CC   antimetastatic, anti-angiogenic, antimalarial and modulates adhesion,
CC   proliferation and mobility of cells.
XX
SQ   Sequence   331 AA;


  Query Match             100.0%;  Score 1756;  DB 20;  Length 331;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-164;
  Matches  331;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQY 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 menpspaaalgkalcalllatlgaagqplggesicsaralakysitftgkwsqtafpkqy 60

Qy      61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 plfrppaqwssllgaahssdysmwrknqyvsnglrdfaergeawalmkeieaagealqsv 120

Qy     121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 havfsapavpsgtgqtsaelevqrrhslvsfvvrivpspdwfvgvdsldlcdgdrwreqa 180

Qy     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aldlypydagtdsgftfsspnfatipqdtvteitssspshpansfyyprlkalppiarvt 240

Qy     241 LVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 lvrlrqsprafippapvlpsrdneivdsasvpetpldcevslwsswglcgghcgrlgtks 300

Qy     301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
           |||||||||||||||||||||||||||||||
Db     301 rtryvrvqpanngspcpeleeeaecvpdncv 331


RESULT   2
AAW23663
ID   AAW23663 standard; Protein; 331 AA.
XX
AC   AAW23663;
XX
DT   08-MAR-1998  (first entry)
XX
DE   Human neuronal attachment factor-1.
XX
KW   Neuronal attachment factor-1; NAF-1; human; cell-cell interaction;
KW   cell adhesion; spinal cord injury; peripheral nerve damage; tumour;
KW   metastasis; therapy; malaria; wound healing.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT                   /label= Sig_peptide
FT   Protein         24..331
FT                   /label= Mat_protein
XX
PN   WO9729189-A1.
XX
PD   14-AUG-1997.
XX
PF   12-FEB-1996;   96WO-US01857.
XX
PR   12-FEB-1996;   96WO-US01857.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Hastings GA;
XX
DR   WPI; 1997-415344/38.
DR   N-PSDB; AAT78360.
XX
PT   New isolated human neuronal attachment factor-1 - promotes cell-cell
PT   interaction and cell adhesion, used for treating e.g. spinal cord
PT   injury, peripheral nerve damage or tumours
XX
PS   Claim 1; Page 42-43; 55pp; English.
XX
CC   This protein comprises human neuronal attachment factor-1 (NAF-1),
CC   a novel protein that promotes cell-cell interaction and cell
CC   adhesion.  Its amino acid sequence was deduced from a DNA molecule
CC   (see AAT78360) isolated from a cDNA library derived from human
CC   epithelial sarcoma.  NAF-1 shows sequence homology to the rat
CC   F-spondin family.  NAF-1 polypeptides, especially the mature
CC   protein, can be expressed in host cells.  It can be used e.g. to
CC   treat spinal cord injuries or damage to peripheral nerves by
CC   promoting neural cell adhesion and neurite extension, to inhibit
CC   tumour cell metastases, inhibit endothelial cell proliferation,
CC   adhesion and motility, to decrease tumour neovascularisation, to
CC   be angiostatic for tumours, to promote wound healing, and to
CC   modulate haemostasis.  It can also be used to screen for agonists
CC   and for antagonists useful e.g. in the prevention of malaria.
XX
SQ   Sequence   331 AA;


  Query Match             99.5%;  Score 1748;  DB 18;  Length 331;
  Best Local Similarity   99.7%;  Pred. No. 4.3e-163;
  Matches  330;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQY 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 menpspaaalgkalcalllatlgaagqplggesicsaralakysitftgkwsqtafpkqy 60

Qy      61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 plfrppaqwssllgaahssdysmwrknqyvsnglrdfaergeawalmkeieaagealqsv 120

Qy     121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 havfsapavpsgtgqtsaelevqrrhslvsfvvrivpspdwfvgvdsldlcdgdrwreqa 180

Qy     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aldlypydagtdsgftfsspnfatipqdtvteitssspshpansfyyprlkalppiarvt 240

Qy     241 LVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     241 lvrlrqsprafippapvlasrdneivdsasvpetpldcevslwsswglcgghcgrlgtks 300

Qy     301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
```

```
                 ||||||||||||||||||||||||||||||
Db     301 rtryvrvqpanngspcpeleeeaecvpdncv 331


RESULT   3
AAW83328
ID   AAW83328 standard; Protein; 331 AA.
XX
AC   AAW83328;
XX
DT   01-MAR-1999  (first entry)
XX
DE   Human mindin polypeptide.
XX
KW   Mindin; human; integrin ligand; angiogenic disease; cancer;
KW   metastasis; chronic inflammatory disorder; rheumatoid arthritis;
KW   atherosclerosis; macular degeneration, diabetic retinopathy;
KW   restenosis; Alzheimer's disease; neural disorder;
KW   tissue remodelling; therapy; diagnosis.
XX
OS   Homo sapiens.
XX
PN   WO9850073-A1.
XX
PD   12-NOV-1998.
XX
PF   07-MAY-1998;   98WO-US09476.
XX
PR   09-MAY-1997;   97US-0046106.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Jonak ZL,  Lane PA,  Trulli SH,  Tsui P;
XX
DR   WPI; 1999-034688/03.
DR   N-PSDB; AAV72537.
XX
PT   A new polypeptide is the integrin ligand, human mindin - useful in
PT   therapy and diagnostic assays for diseases such as those associated
PT   with angiogenesis     .
XX
PS   Claim 1; Page 29; 39pp; English.
XX
CC   This is the amino acid sequence of human mindin, a novel integrin
CC   ligand.  It has homology to F spondin, and may have a
CC   multifunctional activity in normal and disease states.  The
CC   invention provides human mindin polypeptides and polynucleotides
CC   (see also AAV72537-38) and methods for producing such polypeptides by
CC   recombinant techniques.  It also relates to methods for using such
CC   polypeptides and polynucleotides in the treatment of angiogenic
CC   diseases (cancer, cancer metastasis, chronic inflammatory disorders,
CC   rheumatoid arthritis, atherosclerosis, macular degeneration,
CC   diabetic retinopathy), restenosis, Alzheimer's disease, neural
CC   disorders and tissue remodelling.  The invention also relates to
CC   methods for identifying agonists and antagonists/inhibitors and for
CC   treating conditions associated with human mindin imbalance with the
CC   identified compounds.  Diagnostic assays for detecting diseases
CC   associated with inappropriate human mindin activity or levels are
CC   also claimed.
XX
SQ   Sequence   331 AA;


  Query Match             99.5%;  Score 1747;  DB 20;  Length 331;
  Best Local Similarity   99.4%;  Pred. No. 5.4e-163;
  Matches  329;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQY 60
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       1 menpspaaalgkalcalllatlgaagqplggesicsaralakysitftgnwsqtafpkqy 60

Qy      61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120
           ||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db      61 plfrppaqwssllgaahssdysmwrknqyvsnglrdfaergeawalikeieaagealqsv 120

Qy     121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 havfsapavpsgtgqtsaelevqrrhslvsfvvrivpspdwfvgvdsldlcdgdrwreqa 180

Qy     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aldlypydagtdsgftfsspnfatipqdtvteitssspshpansfyyprlkalppiarvt 240

Qy     241 LVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 lvrlrqsprafippapvlpsrdneivdsasvpetpldcevslwsswglcgghcgrlgtks 300

Qy     301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
           |||||||||||||||||||||||||||||||
Db     301 rtryvrvqpanngspcpeleeeaecvpdncv 331


RESULT   4
AAM38872
ID   AAM38872 standard; Protein; 331 AA.
XX
AC   AAM38872;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 2017.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI58028.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 3; SEQ ID NO 2017; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
```

```
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   331 AA;

  Query Match            99.4%;  Score 1746;  DB 22;  Length 331;
  Best Local Similarity  99.4%;  Pred. No. 6.8e-163;
  Matches  329;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQY 60
           |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db       1 menpspaaalgkalcalllatlgaagqplggesicsarapakysitftgkwsqtafpkqy 60

Qy      61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 plfrppaqwssllgaahssdysmwrknqyvsnglrdfaergeawalmkeieaagealqsv 120

Qy     121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 havfsapavpsgtgqtsaelevqrrhslvsfvvrivpspdwfvgvdsldlcdgdrwreqa 180

Qy     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aldlypydagtdsgftfsspnfatipqdtvteitssspshpansfyyprlkalppiarvt 240

Qy     241 LVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300
           |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 llrlrqsprafippapvlpsrdneivdsasvpetpldcevslwsswglcgghcgrlgtks 300

Qy     301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
           |||||||||||||||||||||||||||||||
Db     301 rtryvrvqpanngspcpeleeeaecvpdncv 331


RESULT   5
AAM40658
ID   AAM40658 standard; Protein; 444 AA.
XX
AC   AAM40658;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 5589.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI59814.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 2; SEQ ID NO 5589; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   444 AA;

  Query Match            99.4%;  Score 1746;  DB 22;  Length 444;
  Best Local Similarity  99.4%;  Pred. No. 1e-162;
  Matches  329;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQY 60
           |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db     114 menpspaaalgkalcalllatlgaagqplggesicsarapakysitftgkwsqtafpkqy 173

Qy      61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     174 plfrppaqwssllgaahssdysmwrknqyvsnglrdfaergeawalmkeieaagealqsv 233

Qy     121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     234 havfsapavpsgtgqtsaelevqrrhslvsfvvrivpspdwfvgvdsldlcdgdrwreqa 293

Qy     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     294 aldlypydagtdsgftfsspnfatipqdtvteitssspshpansfyyprlkalppiarvt 353

Qy     241 LVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300
           |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     354 llrlrqsprafippapvlpsrdneivdsasvpetpldcevslwsswglcgghcgrlgtks 413

Qy     301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
           |||||||||||||||||||||||||||||||
Db     414 rtryvrvqpanngspcpeleeeaecvpdncv 444


RESULT   6
AAY79561
ID   AAY79561 standard; Protein; 331 AA.
XX
AC   AAY79561;
XX
```

DT 15-AUG-2000 (first entry)
XX
DE Cancer specific protein Pro108 useful as prostate cancer marker.
XX
KW Cancer specific gene; CSG; prostate cancer; diagnosis; monitoring;
KW staging; imaging; metastasis; therapy; human; Pro108.
XX
OS Homo sapiens.
XX
PN WO200023108-A1.
XX
PD 27-APR-2000.
XX
PF 18-OCT-1999; 99WO-US23764.
XX
PR 19-OCT-1998; 98US-0104741.
XX
PA (DIAD-) DIADEXUS LLC.
XX
Ali SM, Sun Y, Salceda S, Recipon H, Cafferkey R;

WPI; 2000-339528/29.
DR N-PSDB; AAZ95034.
XX
PT Diagnosing, detecting, staging, monitoring, imaging and treating
PT cancers, especially useful for detecting prostate cancer comprises
PT measuring changes in levels of cancer specific genes in cells, tissues
PT and body fluids -
XX
PS Claim 6; Page 29-30; 35pp; English.
XX
CC The present sequence is that of the protein encoded by human cancer
CC specific gene (CSG) pro108 (see AAZ95034). The CSG was identified
CC using a suppression subtractive hybridization method. It is a
CC diagnostic marker for prostate cancer. In 11 different healthy
CC tissues examined, pro108 mRNA expression was highest in the ovary
CC and uterus. Expression in healthy prostate was low. However,
CC overexpression of CSG pro108 was observed in 13 of 13 prostate
CC cancer tissues examined. CSGs comprising pro108, the encoded
CC polypeptide, or an antibody that specifically binds CSG, are used
CC in claimed methods for the diagnosis, detection, staging,
CC monitoring, imaging and treatment of prostate cancer. The new
CC methods provide earlier diagnosis for the presence and metastasis
CC of prostate cancer, and can be used to determine if a cancer has
CC metastasized, or to monitor the progress or stage of the disease
CC when it has not metastasized.
XX
Sequence 331 AA;

Query Match 99.3%; Score 1744; DB 21; Length 331;
Best Local Similarity 99.4%; Pred. No. 1.1e-162;
Matches 329; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQY 60
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1 menpspaaalgkalcalllatlgaagqplggesicsarapakysitftgkwsqtafpkqy 60

Qy      61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 plfrppaqwssllgaahssdysmwrknqyvsnglrdfaergeawalmkeieaagealqsv 120

Qy     121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 hevfsapavpsgtgqtsaelevqrrhslvsfvvrivpspdwfvgvdsldlcdgdrwreqa 180

Qy     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aldlypydagtdsgftfsspnfatipqdtvteitssspshpansfyyprlkalppiarvt 240

Qy     241 LVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 lvrlrqsprafippapvlpsrdneivdsasvpetpldcevslwsswglcgghcgrlgtks 300
```

Qy 301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
||||||||||||||||||||||||||||||||
Db 301 rtryvrvqpanngspcpeleeeaecvpdncv 331

RESULT 7
AAB82472
ID AAB82472 standard; Protein; 331 AA.
XX
AC AAB82472;
XX
DT 22-AUG-2001 (first entry)
XX
DE Human extracellular matrix protein RG1.
XX
KW RG1; human; extracellular matrix protein; prostate cancer;
KW metastasis; tumour; benign prostatic hyperplasia; therapy;
KW diagnosis; antitumour.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 31..103
FT /label= F-spondin_domain
FT Domain 138..221
FT /label= F-spondin_domain
FT Domain 278..300
FT /label= Thrombospondin_domain
FT Peptide 28..46
FT /label= Immunogenic_peptide
FT /note= "specifically claimed in claim 25"
FT Peptide 77.91
FT /label= Immunogenic_peptide
FT /note= "specifically claimed in claim 26"
FT Peptide 188..210
FT /label= Immunogenic_peptide
FT /note= "specifically claimed in claim 27"
FT Peptide 263..274
FT /label= Immunogenic_peptide
FT /note= "specifically claimed in claim 28"
XX
PN WO200144291-A2.
XX
PD 21-JUN-2001.
XX
PF 15-DEC-2000; 2000WO-US33901.
XX
PR 16-DEC-1999; 99US-0172370.
PR 07-DEC-2000; 2000US-9966561.
XX
PA (SCHD) SCHERING AG.
XX
PI Harkins R, Parkes D, Parry G, Schneider DW, Steinbrecher R;
XX
DR WPI; 2001-398128/42.
DR N-PSDB; AAF90566, AAF90567.
XX
PT Novel human extracellular matrix polypeptide, RG1, useful in research,
PT diagnosis and treatment of metastasis such as prostate cancer -
XX
PS Claim 1; Fig 2; 69pp; English.
XX
CC The present sequence is that of human RG1, a new homologue of the
CC extracellular matrix protein family, which is expressed in prostate
CC tissue and may be over-expressed in prostate tumours. It shows
CC homology to mindin and f-spondin proteins. This protein family is
CC distinguished by 2 conserved spondin domains near the N-terminus
CC and at least 1 thrombospondin type 1 repeat at the C-terminus. The
CC rg1 gene (see AAF90566) was isolated from a database screening. The
CC invention provides human RG polypeptides (especially corresponding
CC to amino acids 1-331, 21-331, 27-331, 28-46 and 77-91 of RG1),
CC polynucleotides encoding them, methods for producing the

```
CC   polypeptides, expression vectors, genetically engineered host cells
CC   for expression of the polypeptides, and antibodies which
CC   specifically bind RG1 or its immunologically active fragment
CC   (especially a polypeptide comprising amino acid 28-46, 77-91,
CC   188-210, or 263-274 of RG1). It also provides methods for utilizing
CC   the polynucleotides, polypeptides and antibodies in research,
CC   diagnosis and therapeutic applications. Thus, the antibodies can
CC   be used for diagnosis and/or detection of RG1 expression, or
CC   can be conjugated to a therapeutic agent, such as a cytotoxic
CC   agent, for administration to patient for treatment of diseases
CC   characterised by RG1 activity or expression, such as prostate
CC   cancer.
XX
SQ   Sequence   331 AA;

  Query Match             99.2%;  Score 1742;  DB 22;  Length 331;
  Best Local Similarity   99.4%;  Pred. No. 1.7e-162;
  Matches  329;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQY 60
           ||||||||||||||||||||||||||||||||||| | ||||||||||||||||||||
Db       1 menpspaaalgkalcalllatlgaagqplggesicsagapakysitftgkwsqtafpkqy 60

Qy      61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 plfrppaqwssllgaahssdysmwrknqyvsnglrdfaergeawalmkeieaagealqsv 120

Qy     121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 havfsapavpsgtgqtsaelevqrrhslvsfvvrivpspdwfvgvdsldlcdgdrwreqa 180

Qy     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aldlypydagtdsgftfsspnfatipqdtvteitssspshpansfyyprlkalppiarvt 240

Qy     241 LVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 lvrlrqsprafippapvlpsrdneivdsasvpetpldcevslwsswglcgghcgrlgtks 300

Qy     301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
           |||||||||||||||||||||||||||||||
Db     301 rtryvrvqpanngspcpeleeeaecvpdncv 331


RESULT   8
AAW70589
ID   AAW70589 standard; Protein; 331 AA.
XX
AC   AAW70589;
XX
DT   21-JAN-1999  (first entry)
XX
DE   Adhesion-modulating protein zsig25.
XX
KW   zsig25; adhesion-modulating protein; prostate cell; prostatic carcinoma;
KW   B-cell cancer; infertility; Wolf-Hirschhorn syndrome;
KW   chromosome 4 (p16.3).
XX
OS   Homo sapiens.
XX
PN   WO9845442-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US07117.
XX
PR   11-JUN-1997;   97US-0049288.
PR   10-APR-1997;   97US-0043421.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Sheppard PO;
XX
DR   WPI; 1998-557522/47.
DR   N-PSDB; AAV63241.
XX
PT   New zsig25 protein and related nucleic acid, fusion proteins,
PT   vectors, transformed cells - and antibodies, involved in modulation
PT   of adhesion, used for diagnosis and treatment of prostatic and
PT   B-cell tumours, stimulation of haematopoietic cells, treatment of
PT   immune deficiency etc.
XX
PS   Claim 7; Page 111; 161pp; English.
XX
CC   The present sequence represents a protein designated zsig25. The zsig25
CC   protein is an adhesion-modulating protein expressed at very high level
CC   in prostate cells. The protein is useful as a diagnostic marker for
CC   prostatic carcinoma and B-cell cancers, possibly also for infertility,
CC   and as a reagent for separating cancerous and non-cancerous cells.
CC   The products may also be used to diagnose or treat Wolf-Hirschhorn
CC   syndrome, associated with a deletion in the region of chromosome 4
CC   (p16.3) where the zsig25 gene is located.
XX
SQ   Sequence   331 AA;

  Query Match             99.1%;  Score 1741;  DB 19;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 2.1e-162;
  Matches  328;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQY 60
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1 menpspaaalgkalcalllatlgaagqplggesicsarapakysitftgkwsqtafpkqy 60

Qy      61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 plfrppaqwssllgaahssdysmwrknqyvsnglrdfaergeawalmkeieaagealqsv 120

Qy     121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 hevfsapavpsgtgqtsaelevqrrhslvsfvvrivpspdwfvgvdsldlcdgdrwreqa 180

Qy     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aldlypydagtdsgftfsspnfatipqdtvteitssspshpansfyyprlkalppiarvt 240

Qy     241 LVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300
           |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 llrlrqsprafippapvlpsrdneivdsasvpetpldcevslwsswglcgghcgrlgtks 300

Qy     301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
           |||||||||||||||||||||||||||||||
Db     301 rtryvrvqpanngspcpeleeeaecvpdncv 331


RESULT   9
AAY41721
ID   AAY41721 standard; Protein; 331 AA.
XX
AC   AAY41721;
XX
DT   07-DEC-1999  (first entry)
XX
DE   Human PRO866 protein sequence.
XX
KW   Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW   probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW   secreted protein; transmembrane protein.
XX
OS   Homo sapiens.
XX
PN   WO9946281-A2.
XX
PD   16-SEP-1999.
XX
```

PF 08-MAR-1999; 99WO-US05028.
XX
PR 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 08-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 09-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081838.
PR 15-APR-1998; 98US-0081952.
PR 15-APR-1998; 98US-0081955.
PR 21-APR-1998; 98US-0082568.
PR 21-APR-1998; 98US-0082569.
PR 22-APR-1998; 98US-0082700.
PR 22-APR-1998; 98US-0082704.
PR 22-APR-1998; 98US-0082804.
PR 23-APR-1998; 98US-0082767.
PR 23-APR-1998; 98US-0082796.
PR 27-APR-1998; 98US-0083336.
PR 28-APR-1998; 98US-0083322.
PR 29-APR-1998; 98US-0083392.
PR 29-APR-1998; 98US-0083495.
PR 29-APR-1998; 98US-0083496.
PR 29-APR-1998; 98US-0083499.
PR 29-APR-1998; 98US-0083500.
PR 29-APR-1998; 98US-0083545.
PR 29-APR-1998; 98US-0083554.
PR 29-APR-1998; 98US-0083558.
PR 29-APR-1998; 98US-0083559.
PR 30-APR-1998; 98US-0083742.
PR 05-MAY-1998; 98US-0084366.
PR 06-MAY-1998; 98US-0084414.
PR 06-MAY-1998; 98US-0084441.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.
PR 07-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085338.
PR 13-MAY-1998; 98US-0085339.
PR 15-MAY-1998; 98US-0085573.
PR 15-MAY-1998; 98US-0085579.
PR 15-MAY-1998; 98US-0085580.
PR 15-MAY-1998; 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.
PR 15-MAY-1998; 98US-0085700.
PR 15-MAY-1998; 98US-0085704.
PR 18-MAY-1998; 98US-0086023.
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086486.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.
XX
PA (GETH) GENENTECH INC.
XX
PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
XX
DR WPI; 1999-551358/46.
DR N-PSDB; AAZ34089.
XX
PT New secreted and transmembrane polypeptides and their polynucleotides,
PT useful for treating blood coagulation disorders, cancers and cellular
PT adhesion disorders -
XX
PS Claim 12; Fig 87; 530pp; English.
XX
CC The present invention describes secreted and transmembrane polypeptides
CC and their polynucleotides. The nucleotide sequences are useful as
CC sources of probes, primers, for chromosome mapping, and for generation
CC of antisense sequences. They can also be used to create transgenic
CC animals. The proteins can be used to treat a variety of diseases and
CC disorders, depending on their function. Diseases that may be treated
CC include blood coagulation disorders, cancers and cellular adhesion
CC disorders. They may also be used to raise antibodies. AAZ33891 to
CC AAZ34338, and AAY41685 to AAY41774 represent polynucleotide and
CC polypeptide sequence given in the exemplification of the present
CC invention.
XX
SQ Sequence 331 AA;

Query Match 99.1%; Score 1741; DB 20; Length 331;
Best Local Similarity 99.1%; Pred. No. 2.1e-162;
Matches 328; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQY 60
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1 menpspaaalgkalcalllatlgaagqplggesicsarapakysitftgkwsqtafpkqy 60

Qy      61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 plfrppaqwssllgaahssdysmwrknqyvsnglrdfaergeawalmkeieaagealqsv 120

Qy     121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 hevfsapavpsgtgqtsaelevqrrhslvsfvvrivpspdwfvgvdsldlcdgdrwreqa 180

Qy     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aldlypydagtdsgftfsspnfatipqdtvteitssspshpansfyyprlkalppiarvt 240

Qy     241 LVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300
           |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 llrlrqsprafippapvlpsrdneivdsasvpetpldcevslwsswglcgghcgrlgtks 300

Qy     301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
           ||||||||||||||||||||||||||||||
Db     301 rtryvrvqpanngspcpeleeeaecvpdncv 331
```

```
RESULT  10
AAB44277
ID   AAB44277 standard; Protein; 331 AA.
XX
AC   AAB44277;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human PRO866 (UNQ435) protein sequence SEQ ID NO:236.
XX
KW   Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
KW   expressed sequence tag; detection; cancer.
XX
OS   Homo sapiens.
XX
PN   WO200053756-A2.
XX
PD   14-SEP-2000.
XX
PF   18-FEB-2000; 2000WO-US04341.
XX
PR   08-MAR-1999;   99WO-US05028.
PR   12-MAR-1999;   99US-0123957.
PR   29-MAR-1999;   99US-0126773.
PR   21-APR-1999;   99US-0130232.
PR   28-APR-1999;   99US-0131445.
PR   14-MAY-1999;   99US-0134287.
PR   23-JUN-1999;   99US-0141037.
PR   26-JUL-1999;   99US-0145698.
PR   29-OCT-1999;   99US-0162506.
PR   30-NOV-1999;   99WO-US28313.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28565.
PR   16-DEC-1999;   99WO-US30095.
PR   30-DEC-1999;   99WO-US31243.
PR   30-DEC-1999;   99WO-US31274.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Botstein D,  Desnoyers L,  Eaton DL;
PI   Ferrara N,  Filvaroff E,  Fong S,  Gao W,  Gerber H,  Gerritsen ME;
PI   Goddard A,  Godowski PJ,  Grimaldi CJ,  Gurney AL,  Hillan KJ;
PI   Kljavin IJ,  Kuo SS,  Napier MA,  Pan J,  Paoni NF,  Roy MA;
PI   Shelton DL,  Stewart TA,  Tumas D,  Williams PM,  Wood WI;
XX
DR   WPI; 2000-611443/58.
DR   N-PSDB; AAC78507.
XX
PT   Novel PRO polypeptides and polynucleotides used in detection methods,
PT   to target bioactive molecules to specific cells, and to modulate
PT   cellular activities -
XX
PS   Claim 12; Fig 87; 636pp; English.
XX
CC   AAC78458 to AAC78599 represent polynucleotide and EST (expressed
CC   sequence tag) sequences which encode secreted or transmembrane PRO
CC   polypeptides. The PRO polynucleotides and polypeptides have cytostatic
CC   activity. The polynucleotides and polypeptides can be used for detecting
CC   the presence of PRO polypeptides in samples, for linking bioactive
CC   molecules to cells and for modulating biological activities of cells,
CC   using the polypeptides for specific targeting. The polypeptide targeting
CC   can be used to kill the target cells, e.g. for the treatment of cancers.
CC   The polypeptide pairs provide specific targeting of bioactive molecules
CC   to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
CC   the isolation of the PRO polynucleotide sequences.
XX
SQ   Sequence   331 AA;
```

```
  Query Match           99.1%;  Score 1741;  DB 21;  Length 331;
  Best Local Similarity 99.1%;  Pred. No. 2.1e-162;
  Matches  328;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQY 60
           ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1 menpspaaalgkalcalllatlgaagqplggesicsarapakysitftgkwsqtafpkqy 60

Qy      61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 plfrppaqwssllgaahssdysmwrknqyvsnglrdfaergeawalmkeieaagealqsv 120

Qy     121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 hevfsapavpsgtgqtsaelevqrrhslvsfvvrivpspdwfvgvdsldlcdgdrwreqa 180

Qy     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aldlypydagtdsgftfsspnfatipqdtvteitssspshpansfyyprlkalppiarvt 240

Qy     241 LVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300
           |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 llrlrqsprafippapvlpsrdneivdsasvpetpldcevslwsswglcgghcgrlgtks 300

Qy     301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
           |||||||||||||||||||||||||||||||
Db     301 rtryvrvqpanngspcpeleeeaecvpdncv 331
```

```
RESULT  11
AAB33465
ID   AAB33465 standard; Protein; 331 AA.
XX
AC   AAB33465;
XX
DT   29-JAN-2001  (first entry)
XX
DE   Human PRO866 protein UNQ435 SEQ ID NO:258.
XX
KW   Human; immune related disease; diagnosis; antiinflammatory; cardiant;
KW   dermatological; antiarthritic; antirheumatic; immunosuppressive;
KW   haemostatic; antithyroid; antidiabetic; nootropic; neuroprotective;
KW   antianaemic; hepatotropic; virucide; antipsoriatic; antiallergic;
KW   antiasthmatic; systemic lupus erythematosus; rheumatoid arthritis;
KW   osteoarthritis; spondyloarthropathy; systemic sclerosis; sarcoidosis;
KW   idiopathic inflammatory myopathy; Sjogren's syndrome; thyroiditis;
KW   systemic vasculitis; autoimmune haemolytic anaemia; diabetes mellitus;
KW   autoimmune thrombocytopaenia; immune-mediated renal disease;
KW   demyelinating disease; hepatobiliary disease; Whipple's disease;
KW   inflammatory bowel disease; gluten-sensitive enteropathy;
KW   autoimmune disease; immune-mediated skin disease; allergic disease;
KW   immunological disease; transplantation associated disease;
KW   graft rejection; graft-versus-host-disease.
XX
OS   Homo sapiens.
XX
PN   WO200053758-A2.
XX
PD   14-SEP-2000.
XX
PF   02-MAR-2000; 2000WO-US05841.
XX
PR   08-MAR-1999;   99WO-US05028.
PR   10-MAR-1999;   99US-0123618.
PR   12-MAR-1999;   99US-0123957.
PR   23-MAR-1999;   99US-0125775.
PR   12-APR-1999;   99US-0128849.
PR   20-APR-1999;   99WO-US08615.
PR   28-APR-1999;   99US-0131445.
PR   04-MAY-1999;   99US-0132371.
PR   14-MAY-1999;   99US-0134287.
PR   02-JUN-1999;   99WO-US12252.
```

```
PR   23-JUN-1999;   99US-0141037.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   28-JUL-1999;   99US-0146222.
PR   01-SEP-1999;   99WO-US20111.
PR   08-SEP-1999;   99WO-US20594.
PR   13-SEP-1999;   99WO-US20944.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   05-OCT-1999;   99WO-US23089.
PR   29-OCT-1999;   99US-0162506.
PR   29-NOV-1999;   99WO-US28214.
PR   30-NOV-1999;   99WO-US28313.
PR   30-NOV-1999;   99WO-US28409.
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28564.
PR   02-DEC-1999;   99WO-US28565.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30999.
PR   30-DEC-1999;   99WO-US31274.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Goddard A,  Gurney AL,  Hebert C,  Henzel W;
PI   Kabakoff RC,  Lu Y,  Pan J,  Pennica D,  Shelton DL,  Smith V;
PI   Stewart TA,  Tumas D,  Watanabe CK,  Wood WI,  Yan M;
XX
DR   WPI; 2000-572271/53.
DR   N-PSDB; AAC58630.
XX
PT   Sixty four PRO polypeptides, useful in the diagnosis and treatment of
PT   immune related disorders, e.g. systemic lupus erythematosis, rheumatoid
PT   arthritis, osteoarthritis, thyroiditis and diabetes mellitus -
XX
PS   Claim 33; Fig 104; 309pp; English.
XX
CC   The present invention describes sixty four human PRO proteins which can
CC   be used in the treatment of immune related diseases. The human PRO
CC   proteins, anti-PRO antibodies, agonists and antagonists are useful for
CC   treating and diagnosing immune related disorders. The disorders are
CC   selected from systemic lupus erythematosus, rheumatoid arthritis,
CC   osteoarthritis, juvenile chronic arthritis, spondyloarthropathies,
CC   systemic sclerosis, idiopathic inflammatory myopathies, Sjogren's
CC   syndrome, systemic vasculitis, sarcoidosis, autoimmune haemolytic
CC   anaemia, autoimmune thrombocytopaenia, thyroiditis, diabetes mellitus,
CC   immune-mediated renal disease, demyelinating diseases of the central
CC   and peripheral nervous systems, hepatobiliary diseases, inflammatory
CC   bowel disease, gluten-sensitive enteropathy and Whipple's disease,
CC   autoimmune or immune-mediated skin diseases, allergic diseases,
CC   immunological diseases of the lung, and transplantation associated
CC   diseases including graft rejection and graft-versus-host-disease.
CC   AAC58397 to AAC58578 represent PCR primers and hybridisation probes used
CC   in the isolation of human PRO sequences. AAC58579 to AAC58642 and
CC   AAB33414 to AAB33477 represent human PRO polynucleotide and protein
CC   sequences given in the exemplification of the present invention.
XX
SQ   Sequence   331 AA;

  Query Match             99.1%;  Score 1741;  DB 21;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 2.1e-162;
  Matches  328;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQY 60
                                                                        
Db       1 menpspaaalgkalcalllatlgaagqplggesicsarapakysitftgkwsqtafpkqy 60

Qy      61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120
                                                                        
Db      61 plfrppaqwssllgaahssdysmwrknqyvsnglrdfaergeawalmkeieaagealqsv 120

Qy     121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180
                                                                        
Db     121 hevfsapavpsgtgqtsaelevqrrhslvsfvvrivpspdwfvgvdsldlcdgdrwreqa 180

Qy     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240
                                                                        
Db     181 aldlypydagtdsgftfsspnfatipqdtvteitssspshpansfyyprlkalppiarvt 240

Qy     241 LVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300
                                                                        
Db     241 llrlrqsprafippapvlpsrdneivdsasvpetpldcevslwsswglcgghcgrlgtks 300

Qy     301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
                                            
Db     301 rtryvrvqpanngspcpeleeeaecvpdncv 331


RESULT  12
AAY95349
ID   AAY95349 standard; Protein; 331 AA.
XX
AC   AAY95349;
XX
DT   25-SEP-2000  (first entry)
XX
DE   Human PRO866 antitumour protein.
XX
KW   PRO866; human; antitumour; tumour; therapy; cytostatic;
KW   breast cancer; ovarian cancer; renal cancer; colorectal cancer;
KW   uterine cancer; prostate cancer; lung cancer; bladder cancer;
KW   central nervous system cancer; melanoma; leukaemia; neoplasm.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide          1..26
FT                    /label= Signal_peptide
FT   Protein          27..331
FT                    /label= PRO866
FT   Modified-site    26..32
FT                    /note= "N-myristoylation"
FT   Modified-site    74..80
FT                    /note= "N-myristoylation"
FT   Modified-site    132..138
FT                    /note= "N-myristoylation"
FT   Modified-site    134..140
FT                    /note= "N-myristoylation"
FT   Modified-site    190..196
FT                    /note= "N-myristoylation"
FT   Modified-site    287..293
FT                    /note= "N-myristoylation"
FT   Modified-site    290..296
FT                    /note= "N-myristoylation"
FT   Modified-site    144..148
FT                    /note= "protein kinase phosphorylation site"
FT   Region           131..135
FT                    /note= "glycosaminoglycan attachment site"
XX
PN   WO200037638-A2.
XX
PD   29-JUN-2000.
XX
PF   02-DEC-1999;   99WO-US28565.
XX
PR   22-DEC-1998;   98US-0113296.
PR   08-MAR-1999;   99WO-US05028.
```

```
PR   21-APR-1999;   99US-0130232.
PR   28-APR-1999;   99US-0131445.
PR   14-MAY-1999;   99US-0134287.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Goddard A,  Godowski PJ,  Gurney AL,  Marsters SA;
PI   Napier MA,  Pitti RM,  Wood WI;
XX
DR   WPI; 2000-442668/38.
DR   N-PSDB; AAA49728.
XX
PT   Novel composition to inhibit neoplastic cell growth or for treating
PT   tumor in mammal comprises polypeptides PRO179, PRO207, PRO320, PRO219,
PT   PRO221, PRO224, PRO328, PRO301, PRO526, PRO362, PRO356, PRO509 or
PT   PRO866  -
XX
PS   Claim 19; Fig 26; 172pp; English.
XX
CC   The present sequence is that of human antitumour protein PRO866,
CC   as deduced from a foetal kidney cDNA clone (see AAA49728).  PRO866
CC   shows homology to members of the mindin/spondin family of proteins.
CC   A claimed method for inhibiting the growth of a tumour cell
CC   comprises exposing the tumor cell to PRO179, PRO207, PRO320,
CC   PRO219, PRO221, PRO224, PRO328, PRO301, PRO526, PRO362, PRO356,
CC   PRO509 or PRO866 (see AAY95337-49), their agonists or chimeric
CC   polypeptides incorporating them.  The tumour is especially a cancer
CC   selected from breast, ovarian, renal, colorectal, uterine,
CC   prostate, lung, bladder and central nervous system cancer, melanoma
CC   and leukaemia.  Methods for the recombinant expression of the
CC   antitumour proteins are also provided.
XX
SQ   Sequence   331 AA;

  Query Match             99.1%;  Score 1741;  DB 21;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 2.1e-162;
  Matches  328;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQY 60
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1 menpspaaalgkalcalllatlgaagqplggesicsarapakysitftgkwsqtafpkqy 60

Qy      61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 plfrppaqwssllgaahssdysmwrknqyvsnglrdfaergeawalmkeieaagealqsv 120

Qy     121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 hevfsapavpsgtgqtsaelevqrrhslvsfvvrivpspdwfvgvdsldlcdgdrwreqa 180

Qy     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aldlypydagtdsgftfsspnfatipqdtvteitssspshpansfyyprlkalppiarvt 240

Qy     241 LVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300
           |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 llrlrqsprafippapvlpsrdneivdsasvpetpldcevslwsswglcgghcgrlgtks 300

Qy     301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
           |||||||||||||||||||||||||||||||
Db     301 rtryvrvqpanngspcpeleeeaecvpdncv 331


RESULT  13
AAG75576
ID   AAG75576 standard; Protein; 385 AA.
XX
AC   AAG75576;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Human colon cancer antigen protein SEQ ID NO:6340.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US26524.
XX
PR   29-SEP-1999;   99US-0157137.
PR   03-NOV-1999;   99US-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Barash SC,  Birse CE,  Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   N-PSDB; AAH34981.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS   Claim 11; Page 7793-7795; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer-associated Ps,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC   and AAB77789 represent sequences used in the exemplification of the
CC   present invention.
CC   N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ   Sequence   385 AA;

  Query Match             99.1%;  Score 1740;  DB 22;  Length 385;
  Best Local Similarity   99.1%;  Pred. No. 3.2e-162;
  Matches  328;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQY 60
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      55 menpspaaalgkalcalllatlgaagqplggesicsarapakysitftgkwsqtafpkqy 114

Qy      61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     115 plfrppaqwssllgaahssdysmwrknqyvsnglrdfaergeawalmkeieaagealqsv 174

Qy     121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175 hxvfsapavpsgtgqtsaelevqrrhslvsfvvrivpspdwfvgvdsldlcdgdrwreqa 234

Qy     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     235 aldlypydagtdsgftfsspnfatipqdtvteitssspshpansfyyprlkalppiarvt 294
```

```
Qy     241 LVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     295 lxrlrqsprafippapvlpsrdneivdsasvpetpldcevslwsswglcgghcgrlgtks 354

Qy     301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
           |||||||||||||||||||||||||||||||
Db     355 rtryvrvqpanngspcpeleeeaecvpdncv 385


RESULT  14
AAB34753
ID   AAB34753 standard; Protein; 249 AA.
XX
AC   AAB34753;
XX
DT   26-JAN-2001  (first entry)
XX
DE   Human secreted protein fragment encoded by DNA clone vp22 1.

KW   Secreted protein; human; autoimmune disorder; multiple sclerosis; ulcer;
KW   systemic lupus erythematosus; rheumatoid arthritis; anaemia; stroke;
KW   haematopoiesis regulation; tissue regrowth; wound healing; haemophilia;
KW   Alzheimer's disease; Parkinson's disease; Shy-drager syndrome; cancer;
KW   contraceptive; infection; growth inhibition; hyperproliferative disorder;
KW   psoriasis.
XX
OS   Homo sapiens.
XX
PN   WO200055375-A1.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US07285.
XX
PR   17-MAR-1999;   99US-0124808.
PR   17-MAR-1999;   99US-0124916.
PR   17-AUG-1999;   99US-0149639.
PR   01-OCT-1999;   99US-0157247.
PR   29-NOV-1999;   99US-0167824.
PR   15-FEB-2000; 2000US-0182711.
XX
PA   (ALPH-) ALPHAGENE INC.
XX
PI   Valenzuela D,  Yuan O,  Hoffman H,  Hall J,  Rapiejko P;
XX
DR   WPI; 2000-638211/61.
DR   N-PSDB; AAC59794.

PT   Novel proteins and polypeptides useful for the treatment of e.g
PT   multiple sclerosis, systemic lupus erythmatosus, rheumatoid arthritis,
PT   cancer, Alzheimer's disease, Parkinson's disease, stroke, anemia and
PT   ulcers -
XX
PS   Disclosure; Page 477-478; 493pp; English.
XX
CC   This invention relates to 59 human secreted proteins and the nucleotide
CC   sequences encoding them. Sequences AAC59788-C59846 and AAB34687-B34745
CC   represent the proteins and their encoding nucleotide sequences, and
CC   sequences AAB34746-B34771 represent fragments of the proteins. Probes
CC   for the DNA sequences are represented by sequences AAC59847-C59596. The
CC   proteins exhibit neuroprotective, dermatological, immunosuppressive,
CC   antiinflammatory, antianaemic, nootropic, antiparkinsonian,
CC   cerebroprotective, haemostatic, vulnerary, cytostatic, antipsoriatic,
CC   antibacterial, virucide, and fungicide activity. The proteins and
CC   nucleotide sequences are useful as nutritional sources or supplements
CC   and in research. The proteins are useful for treating immune deficiency
CC   and disorders, which may be genetic or resulting from infections,
CC   autoimmune disorders such as multiple sclerosis, systemic lupus
CC   erythmatosus, rheumatoid arthritis, and for treating myeloid or lymphoid
CC   cell deficiencies such as anaemias by regulating haematopoiesis. The
CC   proteins are also useful in compositions for bone, cartilage, tendon,
CC   ligament and/or nerve tissue growth or regeneration, for wound healing,
CC   tissue repair and replacement and in the treatment of wounds, incisions
CC   and ulcers. Other uses include in the treatment of central and
CC   peripheral nervous system and neuropathies such as Alzheimer's and
CC   Parkinson's diseases and Shy-Drager syndrome, and mechanical and
CC   traumatic disorders, such as spinal cord disorders, head trauma and
CC   stroke. The proteins may also be used as a contraceptive, and for
CC   treating coagulation disorders such as haemophilias. The protein and
CC   nucleotide sequences with cadherin activity are useful for treating
CC   cancer. Other uses for the protein include for inhibiting the growth,
CC   infection or function of, or killing, infectious agents such as bacteria,
CC   virus, fungi and other parasites, for effecting bodily characteristics
CC   such as height, weight, hair colour, effecting biorhythms or cardiac
CC   cycles or rhythms, effecting metabolism, catabolism, anabolism,
CC   processing, utilization, storage or elimination of dietary fat, lipid,
CC   protein, carbohydrate, vitamins, minerals, cofactors, effecting
CC   behavioural characteristics, providing analgesic effects and for treating
CC   hyperproliferative disorders such as psoriasis.
XX
SQ   Sequence   249 AA;


  Query Match             75.7%;  Score 1329;  DB 21;  Length 249;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-122;
  Matches  249;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      83 MWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSVHAVFSAPAVPSGTGQTSAELEV 142
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mwrknqyvsnglrdfaergeawalmkeieaagealqsvhavfsapavpsgtgqtsaelev 60

Qy     143 QRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQAALDLYPYDAGTDSGFTFSSPNF 202
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 qrrhslvsfvvrivpspdwfvgvdsldlcdgdrwreqaaldlypydagtdsgftfsspnf 120

Qy     203 ATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVTLVRLRQSPRAFIPPAPVLPSRD 262
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 atipqdtvteitssspshpansfyyprlkalppiarvtlvrlrqsprafippapvlpsrd 180

Qy     263 NEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKSRTRYVRVQPANNGSPCPELEEE 322
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 neivdsasvpetpldcevslwsswglcgghcgrlgtksrtryvrvqpanngspcpeleee 240

Qy     323 AECVPDNCV 331
           |||||||||
Db     241 aecvpdncv 249


RESULT  15
AAY73490
ID   AAY73490 standard; Protein; 299 AA.
XX
AC   AAY73490;
XX
DT   29-FEB-2000  (first entry)
XX
DE   Human secreted protein clone yk143_1 protein sequence SEQ ID NO:202.
XX
KW   Human; secreted protein; immunostimulatory; haemostatic; cytokine;
KW   proliferative; differentiative; chemotactic; chemokinetic; vaccine;
KW   thrombolytic; antiinflammatory; cytostatic; immunosuppressive;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO9958642-A2.
XX
PD   18-NOV-1999.
XX
PF   14-MAY-1999;   99WO-US10843.
XX
PR   14-MAY-1998;   98US-0085472.
PR   17-AUG-1998;   98US-0096824.
PR   11-SEP-1998;   98US-0099843.
PR   11-SEP-1998;   98US-0099950.
```

PR 15-SEP-1998; 98US-0100424.
PR 29-SEP-1998; 98US-0102329.
PR 09-OCT-1998; 98US-0103615.
PR 11-DEC-1998; 98US-0111799.
PR 14-DEC-1998; 98US-0112159.
PR 31-DEC-1998; 98US-0114415.
PR 10-FEB-1999; 99US-0248059.
PR 06-APR-1999; 99US-0287150.
PR 13-MAY-1999; 99US-0311021.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Wong GG, Clark HF, Fechtel K, Agostino MJ;
XX
DR WPI; 2000-053095/04.
DR N-PSDB; AAZ52575.
XX
PT Novel polynucleotides and proteins having biological activities which
PT make them suitable for treating, preventing or ameliorating medical
PT conditions in humans or animals -
XX
Claim 211; Page 710-711; 730pp; English.
XX
CC The present invention describes human secreted proteins encoded by
CC polynucleotides obtained from adult testes, foetal brain, adult brain,
CC brain (foetal and adult), foetal kidney, adult spleen, and adult thymus
CC cDNA libraries. The polynucleotides and proteins are predicted to have
CC biological activities which would make them suitable for treating,
CC preventing or ameliorating medical conditions in humans and animals.
CC Suggested activities include nutritional activity, cytokine and cell
CC proliferation/differentiation activity, immune stimulating (e.g. as
CC vaccines) or suppressing activity, haematopoiesis regulating activity,
CC tissue growth activity, activin/inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, cadherin/tumour invasion
CC suppressor activity, and tumour inhibition activity. The polynucleotides
CC are also stated to be useful for gene therapy. Therapeutic compositions
CC are also presently valuable for veterinary applications. AAZ52475 to
CC AAZ52581 encode human secreted proteins, and AAY73390 to AAY73500
CC represent human secreted proteins, given in the present invention.
XX
SQ Sequence 299 AA;

Query Match 62.7%; Score 1101.5; DB 21; Length 299;
Best Local Similarity 90.3%; Pred. No. 1e-99;
Matches 214; Conservative 5; Mismatches 15; Indels 3; Gaps 2;

```
Qy      83 MWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSVHAVFSAPAVPSGTGQTSAELEV 142
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mwrknqyvsnglrdfaergeawalmkeieaagealqsvhavfsapavpsgtgqtsaelev 60

Qy     143 QRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQAALDLYPYDAGTDSGFTFSSPNF 202
           |||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db      61 qrrhslvsfvvrivpspdwfvgvnsldlcdgdrwreqaaldlypydagtdsgftfsspnf 120

Qy     203 ATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVTLVRLRQSPRAFIPPAPVLPSRD 262
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 atipqdtvteitssspshpansfyyprlkalppiarvtlvrlrqsprafippapvlpsrd 180

Qy     263 NEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKSRTRYVRVQPA--NNGSPCP 317
           |||||||||||||||||||||||||||||       : :   :|  ||    |:| |
Db     181 neivdsasvpetpldcevslwsswglcggplweardqeqds-lrpgparqqtgapap 236
```

Search completed: December 10, 2001, 19:46:03
Job time: 22032 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2001, 19:17:38 ; Search time 16.91 Seconds
(without alignments)
1491.057 Million cell updates/sec

Title: US-09-170-042A-2
Perfect score: 1756
Sequence: 1 MENPSPAAALGKALCALLLA.....NGSPCPELEEEAECVPDNCV 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

al number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	462.5	26.3	807	2	A38152	F-spondin - rat
2	448.5	25.5	803	2	A47723	F-spondin precurso
3	406	23.1	805	2	T34212	hypothetical prote
4	127.5	7.3	741	2	I48694	probable transcrip
5	117	6.7	772	2	A55004	transcription fact
6	108	6.2	123	2	S49108	TRAP-C2 protein -
7	108	6.2	742	2	A49672	transcription fact
8	107	6.1	808	2	T10171	phospholipase D (E
9	104.5	6.0	440	2	T24232	hypothetical prote
10	104	5.9	1251	2	A57293	latent transformin
11	102	5.8	483	1	VCBPI3	minor coat protein
12	100.5	5.7	810	2	D96566	hypothetical prote
13	100.5	5.7	2957	2	T33152	hypothetical prote
14	99.5	5.7	424	2	C70651	hypothetical prote
15	99.5	5.7	590	2	I46687	complement compone
16	99.5	5.7	724	2	A48569	antigen Em100 - Ei
17	99.5	5.7	812	2	T03659	phospholipase D (E
18	97.5	5.6	808	2	T04092	phospholipase D (E
19	97	5.5	534	2	T41081	hypothetical prote
20	97	5.5	1360	2	T33922	hypothetical prote
21	97	5.5	1584	2	T00026	brain-specific ang
22	96.5	5.5	598	2	A57249	beta-galactosidase
23	96.5	5.5	712	2	A45638	immunodominant mic
24	96.5	5.5	812	2	T03402	probable phospholi
25	96	5.5	591	1	C8HUB	complement C8 beta
26	95	5.4	1070	2	S75712	cellulase (EC 3.2.
27	94	5.4	809	2	T11695	phospholipase D (E
28	93	5.3	3027	2	JQ1917	polyprotein - pars
29	92.5	5.3	912	2	A54423	brevican precursor
30	92	5.2	1306	2	S25370	MSB2 protein - yea
31	91.5	5.2	341	2	JC1201	carboxymethylcellu
32	91	5.2	1007	2	T01437	hypothetical prote
33	91	5.2	4976	2	T14165	peptide synthetase
34	90.5	5.2	777	2	T00208	transposase-like p
35	90.5	5.2	1333	2	G84542	probable retroelem
36	90	5.1	497	2	T41015	proline rich prote
37	90	5.1	584	2	S76424	hypothetical prote
38	90	5.1	1444	2	T18856	angiogenesis inhib
39	90	5.1	1666	2	T43169	hypothetical prote
40	90	5.1	2265	1	FNBO	fibronectin - bovi
41	89.5	5.1	919	2	T32541	unc-5 protein - Ca
42	89.5	5.1	947	1	B44294	unc-5 protein, lon
43	89	5.1	668	2	B86831	hypothetical prote
44	89	5.1	1257	2	T09493	period protein hom
45	88	5.0	620	2	A70525	hypothetical prote

ALIGNMENTS

RESULT 1
A38152
F-spondin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 17-Nov-2000
C;Accession: A38152
R;Klar, A.; Baldassare, M.; Jessell, T.M.
Cell 69, 95-110, 1992
A;Title: F-spondin: a gene expressed at high levels in the floor plate encodes a secr
A;Reference number: A38152; MUID:92208952
A;Accession: A38152
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-807 <KLA>
A;Cross-references: GB:M88469; NID:g204176; PIDN:AAA41174.1; PID:g204177
A;Experimental source: embryo floor plate
A;Note: sequence extracted from NCBI backbone (NCBIN:90877, NCBIP:90878)
C;Superfamily: F-spondin; thrombospondin type 1 repeat homology
F;441-495/Domain: thrombospondin type 1 repeat homology <THR2>
F;500-555/Domain: thrombospondin type 1 repeat homology <THR3>
F;557-611/Domain: thrombospondin type 1 repeat homology <THR4>
F;613-666/Domain: thrombospondin type 1 repeat homology <THR1>
F;667-721/Domain: thrombospondin type 1 repeat homology <THR5>
F;753-807/Domain: thrombospondin type 1 repeat homology <THR6>

Query Match 26.3%; Score 462.5; DB 2; Length 807;
Best Local Similarity 34.2%; Pred. No. 2.9e-29;
Matches 116; Conservative 49; Mismatches 137; Indels 37; Gaps 12;

```
Qy       9 ALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQ 68
           :| | ||       |   :|:     | |   ||| :|| | ||:   || ||  |
Db     176 SLTKKLCEQDPTLDGVTDRPI---LDCCACGTAKYRLTFYGNWSEKTHPKDYP--RRANH 230

Qy      69 WSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEI-EAAGEALQSVHAVFSAP 127
           ||:::| :|| :| :|    | | |::  || |    : :|| : : | |  : |    |
Db     231 WSAIIGGSHSKNYVLWEYGGYASEGVKQVAELGSPVKMEEEIRQQSDEVLTVIKAKAQWP 290

Qy     128 --------AVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGD-RWRE 178
                 | |      |||  | |   |:||:  : |||||  ||: : |||  :  | :
Db     291 SWQPVNVRAAP------SAEFSVDRTRHLMSFLTMMGPSPDWNVGLSAEDLCTKECGWVQ 344

Qy     179 QAALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIAR 238
           :   || |:|||||| |: |||  ||||: :  :||    || : || |   ::  :||
Db     345 KVVQDLIPWDAGTDSGVTYESPNKPTIPQEKIRPLTSL--DHPQSPFYDPEGGSITQVAR 402

Qy     239 VTLVRL-RQSPRAFIPPAPVLPSRDNEIVDSASVPE------TPLDCEVSLWSSWGLCGG 291
           | : |: |:  :  | |  |    |: : | |  ||      ||  |  | || |  |
Db     403 VVIERIARKGEQCNIVPDNV----DDIVADLA--PEEKDEDDTPETCIYSNWSPWSACSS 456

Qy     292 HCGRLGTKSRTRYVRVQPANNGSPCPELEEEAECVPDNC 330
             | : | | :: |  :   |||: ::   |:   |
```

```
Db    457 STCEKGKRMRQRMLKAQ-LDLSVPCPDTQDFQPCMGPGC 494


RESULT   2
A47723
F-spondin precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 17-Nov-2000
C;Accession: A47723
R;Ruiz i Altaba, A.; Cox, C.; Jessell, T.M.; Klar, A.
Proc. Natl. Acad. Sci. U.S.A. 90, 8268-8272, 1993
A;Title: Ectopic neural expression of a floor plate marker in frog embryos injected with
A;Reference number: A47723; MUID:93376785
A;Accession: A47723
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-803 <RUI>
A;Cross-references: GB:L09123; NID:g409244; PIDN:AAA19105.1; PID:g409245
C;Superfamily: F-spondin; thrombospondin type 1 repeat homology
F;435-489/Domain: thrombospondin type 1 repeat homology <THR2>
F;607-662/Domain: thrombospondin type 1 repeat homology <THR1>


  Query Match           25.5%;  Score 448.5;  DB 2;  Length 803;
  Best Local Similarity   34.3%;  Pred. No. 3.8e-28;
  Matches  115;  Conservative   48;  Mismatches  141;  Indels   31;  Gaps   12;

Qy      9 ALGKALCALLLATLGAAGQPLGGESI---CSARALAKYSITFTGKWSQTAFPKQYPLFRP 65
          :| | :| | | |   |      | |     | |   ||| :|| | ||: | || ||  |
Db    172 SLTKRMCELDLTLEG------GNEKTIPDCCACGTAKYRLTFYGNWSEKAHPKDYP--RR 223

Qy     66 PAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAG-EALQSVHAVF 124
            ||:::| :|| :| :|   |  |:|::  || |    : :||   | | |  : |
Db    224 ANHWSAIIGGSHSGEYVLWEYGQ-ASDGVKQVAELGSPVKMEEEIRQKGDEVLTVIKAKA 282

Qy    125 SAPAVP--SGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGD-RWREQAA 181
           ||    :     ||| | |   |:||:  : ||||| ||: | |||  :  | ::
Db    283 QWPAWQPLNVRAAPSAEFSVDRSRHLMSFLAMMGPSPDWNVGLTSEDLCTKECGWVQKVV 342

Qy    182 LDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITS-SSPSHPANSFYYPRLKALPPIARVT 240
          || |:|||||| |: |||  ||||| :  :||   |  |: :    |   : |||||
Db    343 QDLIPWDAGTDSGVTYESPNKPTIPQDKIRPLTSLDHPQSPSMT----RGGPIIPIARVV 398

Qy    241 LVRLRQSPRAFIPPAPVLPSRDNEIV-----DSASVPETPLDCEVSLWSSWGLCGGHCGR 295
          : |: :        ::|   ::||    :     :||  |  | || |  |
Db    399 IERIARKGE----QCNIIPDNVDDIVADLVTEEKDEDDTPETCIYSNWSPWSACSSATCD 454

Qy    296 LGTKSRTRYVRVQPANNGSPCPELEEEAECVPDNC 330
          | : | | :: |  :   |||: ::   |:   |
Db    455 KGKRMRQRMLKAQ-LDLSVPCPDTQDFEPCMGPGC 488


RESULT   3
T34212
hypothetical protein F10E7.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34212
R;Pauley, A.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid F10E7.
A;Reference number: Z21489
A;Accession: T34212
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-805 <PAU>
A;Cross-references: EMBL:U41264; PIDN:AAA82427.1; CESP:F10E7.4
C;Genetics:
A;Gene: CESP:F10E7.4
A;Introns: 9/1; 34/3; 57/1; 90/3; 128/3; 162/1; 205/1; 285/2; 417/1; 475/1; 606/1; 745/1


  Query Match           23.1%;  Score 406;  DB 2;  Length 805;
  Best Local Similarity   28.2%;  Pred. No. 9.9e-25;
  Matches  103;  Conservative   53;  Mismatches  131;  Indels   78;  Gaps    9;

Qy     35 CSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGL 94
          | |  :|:| : ||| ||:   || ||       :: :||::|||:||:|     ::|:
Db    177 CCACDIAQYDLEFTGIWSKNTHPKDYPTLEHLTHFTDMLGSSHSSNYSLWTIGGISTDGM 236

Qy     95 RDFAERGEAWALMKEIEAAGEALQSVHAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVR 154
          :: || |  :    | :|    ::|:  |       |   | | ::  | : |  ||
Db    237 KEIAEWGNTYKAEAEAKAKASEVRSLMKV-KGLWFPDVQGTTKSQFVVNKYHHFVSLATM 295

Qy    155 IVPSPDWFVGVDSLDLCDGD-RWREQAALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEI 213
            ||||| ||: |::||  |  | |:   :| |:|||||| |: |||  | |:: :  |
Db    296 FGPSPDWCVGLSSVNLCLPDCTWAEERTFELQPFDAGTDSGPTYMSPNEPTEPREPIHWI 355

Qy    214 TSSSPSHPANSFYYPRLKALPPIARVTL---------------------------------- 241
          |:    :| : ||  :   :| :|:| |
Db    356 TTK--LNPLSPFYNSKSDTIPTLAKVILRRKNVTSSECKSDDILKAEAHNITNTSEDEEY 413

Qy    242 ------------------------VRLRQSPRAFIPPAPVL-----------PSRDNEI 265
                                  :|:|     |   | |             :: ||
Db    414 KDRRECMMTQWEPWSLCSATCGKGIRIRSRVYVFPIKAQVFHCHRQTTEKQFCNAKINEC 473

Qy    266 VDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKSRTRYVRVQPANNGSPCP-ELEEEAE 324
          :| :       |:|| | ||| |   ||  | :|| |   : ||     |  :|| :
Db    474 ENSEAFSS---KCQVSSWGSWGECSVQCGH-GMRSRNR-TFLNPATKSGDCSVDLERKDI 528

Qy    325 CVPDN 329
          || :|
Db    529 CVGEN 533


RESULT   4
I48694
probable transcription factor NFE2L1 - mouse
N;Alternate names: NFE2-related factor 1
C;Species: Mus musculus (house mouse)
C;Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 05-Nov-1999
C;Accession: I48694; S44137
R;McKie, J.; Johnstone, K.; Mattei, M.G.; Scambler, P.
Genomics 25, 716-719, 1995
A;Title: Cloning and mapping of murine Nfe2l1.
A;Reference number: A56006; MUID:95278942
A;Accession: I48694
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-741 <RES>
A;Cross-references: EMBL:X78709; NID:g473089; PIDN:CAA55362.1; PID:g473090
C;Genetics:
A;Gene: NRF1
C;Superfamily: fos/jun DNA-binding domain homology
F;617-658/Domain: fos/jun DNA-binding domain homology <FJD>


  Query Match            7.3%;  Score 127.5;  DB 2;  Length 741;
  Best Local Similarity   22.4%;  Pred. No. 0.021;
  Matches   79;  Conservative   45;  Mismatches  119;  Indels  109;  Gaps   17;

Qy     19 LATLGAAGQPLGGESICSARALAKYSITFTG-KWSQTAFPKQYPLFRPPAQWSSLLGAAH 77
          |  |||   |: |:       | |  |      | |                    |||
Db    153 LEDLGAVAPPVSGD-------LTKEDIDLIDILWRQDID---------------LGAGR 189

Qy     78 S-SDYSMWRKNQYVSNGLRDFAERGEAW------ALMKEIEAAGEALQSVHAVFSA--PA 128
             |||  :| | |   |:|  || : |       || :::   ||  :|  | | |   :
Db    190 EVFDYSHRQKEQDVDKELQDGREREDTWSGEGAEALARDLLVDGETGESFPAQFPADVSS 249

Qy    129 VPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWRE-QAALDLYPY 187
          :|      |    :|  :||:| ::    ||         ||    :|::  : :::
Db    250 IPEAVPSESESPALQ--NSLLSPLLTGTESP--------FDL--EQQWQDLMSIMEMQAM 297

Qy    188 DAGTD-SGFTFSSPNFATIPQDTVTEITSSSPSHPAN-----------------SFYYPR 229
          :  |  |   :::|     | | ::   | :|: | |                   | : |
Db    298 EVNTSASEILYNAP-----PGDPLSSNYSLAPNTPINQNVSLHQASLGGCSQDFSLFSPE 352
```

```
Qy     230 LKALPPIARVTLVRLRQSPRAFIP--------PAPVLPSRDNEIVDSASVPETP------ 275
           :::||  :  ||: |  |    :           |  ||: |   :  | || |
Db     353 VESLPVASSSTLLPLVPSNSTSLNSTFGSTNLAGPFFPSQLNGTANDTSGPELPDPLGGL 412

Qy     276 -----LDCEVSLWSSWGLCGGHCGRLGTKSRTRYVRVQPANNGSPCPELEEE 322
               || |:||                  : ::   |     :||||
Db     413 LDEAMLD-EISLMD-------------------LAIEEGFNPVQASQLEEE 443


RESULT   5
A55004
transcription factor TFC11 - human
N;Alternate names: LCR-F1 protein
C;Species: Homo sapiens (man)
C;Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 05-Nov-1999
C;Accession: A55004; S48097
R;Luna, L.; Johnsen, O.; Skartlien, A.H.; Pedeutour, F.; Turc-Carel, C.; Prydz, H.; Kols
Genomics 22, 553-562, 1994
A;Title: Molecular cloning of a putative novel human bZIP transcription factor on chromo
A;Reference number: A55004; MUID:95095252
A;Accession: A55004
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-772 <LUN>
A;Cross-references: GB:X77366; NID:g541677; PIDN:CAA54555.1; PID:g541678
R;Caterina, J.J.; Donze, D.; Sun, C.W.; Ciavatta, D.J.; Townes, T.M.
Nucleic Acids Res. 22, 2383-2391, 1994
A;Title: Cloning and functional characterization of LCR-F1: a bZIP transcription factor
A;Reference number: S48097; MUID:94310069
A;Accession: S48097
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 326-772 <CAT>
C;Genetics:
A;Gene: GDB:TCF11
A;Cross-references: GDB:293921; OMIM:600115
A;Map position: 17q22-17q22
C;Superfamily: fos/jun DNA-binding domain homology
C;Keywords: DNA binding; leucine zipper; transcription factor
F;648-689/Domain: fos/jun DNA-binding domain homology <FJD>


  Query Match             6.7%;  Score 117;  DB 2;  Length 772;
  Best Local Similarity   23.3%;  Pred. No. 0.16;
  Matches   78;  Conservative   35;  Mismatches  110;  Indels  112;  Gaps   17;

      19 LATLGAAGQPLGGESICSARALAKYSITFTG-KWSQTAFPKQYPLFRPPAQWSSLLGAAH 77
         |  |||    |: |:       | |  |      | |                  |||
     153 LEDLGAVAPPVSGD-------LTKEDIDLIDILWRQDID---------------LGAGR 189

Qy      78 S-SDYSMWRKNQYVSNGLRDFAER----GE-AWALMKEIEAAGEALQSVHAVFSAPAVPS 131
           |||  :| | |   |||  |:    || | || : :   ||  :|  |      |||
Db     190 EVFDYSHRQKEQDVEKELRDGGEQDTWAGEGAEALARNLLVDGETGESFPA-----QVPS 244

Qy     132 GTGQTSAELEVQRR-------------HSLVSFVVRIVPSPD-----------WFVGV 165
           |  ||:  ||   |               : :| :   |||               |
Db     245 GEDQTALSLEECLRLLEATCPFGENAEFPADISSITEAVPSESEPPALQNNLLSPLLTGT 304

Qy     166 DS-LDLCDGDRWRE-QAALDLYPYDAGTD-SGFTFSSPNFATIPQDTVTEITSSSPSHPA 222
           :|  ||    :|::  : :::   :  |  |   :|:|     | | ::   | :|: |
Db     305 ESPFDL--EQQWQDLMSIMEMQAMEVNTSASEILYSAP-----PGDPLSTNYSLAPNTPI 357

Qy     223 NS----------------FYYPRLKALPPIARVTLVRLRQSPRA--------------F 251
           |                : | :::||  :  ||: |  |                  |
Db     358 NQNVSLHQASLGGCSQDFLLFSPEVESLPVASSSTLLPLAPSNSTSLNSTFGSTNLTGLF 417

Qy     252 IPP----------APVLPSRDNEIVDSASVPETPL 276
           ||           | ||     ::| | : |  |
Db     418 FPPQLNGTANDTAGPELPDPLGGLLDEAMLDEISL 452


RESULT   6
S49108
TRAP-C2 protein - Cryptosporidium parvum (fragment)
C;Species: Cryptosporidium parvum
C;Date: 01-Feb-1995 #sequence_revision 12-May-1995 #text_change 09-Sep-1997
C;Accession: S49108
R;Spano, F.S.; Ranucci, L.R.; Catteruccia, F.C.; Saccheo, S.S.; Crisanti, A.C.
submitted to the EMBL Data Library, January 1994
A;Description: Thrombospondin related protein in Cryptosporidium.
A;Reference number: S49108
A;Accession: S49108
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-123 <SPA>
A;Cross-references: EMBL:X77586; NID:g509278; PID:g509279
C;Superfamily: thrombospondin type 1 repeat homology
F;6-61/Domain: thrombospondin type 1 repeat homology <THR1>


  Query Match             6.2%;  Score 108;  DB 2;  Length 123;
  Best Local Similarity   40.4%;  Pred. No. 0.089;
  Matches   23;  Conservative    6;  Mismatches   26;  Indels    2;  Gaps    2;

Qy     275 PLDCEVSLWSSWGLCGGHCGRLGTKSRTRYVRVQPAN-NGSPCPELEEEAECVPDNC 330
           || | || | :|  |   || :| : | | |   | : |   |||  :  ||: | |
Db       5 PLSCTVSEWGNWSRCSLTCG-IGHQMRERSVIKAPKDQNLFQCPETRQIRECIQDTC 60


RESULT   7
A49672
transcription factor Nrf1 - human
N;Alternate names: basic leucine-zipper transcription factor; NF-E2-related factor 1
C;Species: Homo sapiens (man)
C;Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 16-Dec-1998
C;Accession: A49672
R;Chan, J.Y.; Han, X.L.; Kan, Y.W.
Proc. Natl. Acad. Sci. U.S.A. 90, 11371-11375, 1993
A;Title: Cloning of Nrf1, an NF-E2-related transcription factor, by genetic selection
A;Reference number: A49672; MUID:94068605
A;Accession: A49672
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-742 <CHA>
A;Experimental source: erythroleukemia cell line K562
A;Note: sequence extracted from NCBI backbone (NCBIN:140521, NCBIP:140522)
C;Superfamily: fos/jun DNA-binding domain homology
F;618-659/Domain: fos/jun DNA-binding domain homology <FJD>


  Query Match             6.2%;  Score 108;  DB 2;  Length 742;
  Best Local Similarity   23.2%;  Pred. No. 0.78;
  Matches   73;  Conservative   38;  Mismatches  103;  Indels  100;  Gaps   16;

Qy      19 LATLGAAGQPLGGESICSARALAKYSITFTG-KWSQTAFPKQYPLFRPPAQWSSLLGAAH 77
           |  |||    |: |:       | |  |      | |                  |||
Db     153 LEDLGAVAPPVSGD-------LTKEDIDLIDILWRQDID---------------LGAGR 189

Qy      78 S-SDYSMWRKNQYVSNGLRDFAER----GE-AWALMKEIEAAGEALQSVHAVFSA----- 126
           |||  :| | |   |||  |:    || | || : :   ||  :|  | | |
Db     190 EVFDYSHRQKEQDVEKELRDGGEQDTWAGEGAEALARNLLVDGETGESFPAQFPADISSI 249

Qy     127 -PAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWRE-QAALDL 184
            |||| :   : :       ::|:| ::    ||          ||    :|::  : :::
Db     250 TEAVPSESEPPALQ------NNLLSPLLTGTESP--------FDL--EQQWQDLMSIMEM 293

Qy     185 YPYDAGTD-SGFTFSSPNFATIPQDTVTEITSSSPSHPANS----------------FY 226
           :  |  |   :|:|     | | ::   | :|: | |                  :
Db     294 QAMEVNTSASEILYSAP-----PGDPLSTNYSLAPNTPINQNVSLHQASLGGCSQDFLLF 348

Qy     227 YPRLKALPPIARVTLVRLRQSPRA--------------FIPP----------APVLPSRD 262
           | :::||  :  ||: |  |                  | ||           | ||
Db     349 SPEVESLPVASSSTLLPLAPSNSTSLNSTFGSTNLTGLFFPPQLNGTANDTAGPELPDPL 408

Qy     263 NEIVDSASVPETPL 276
```

```
          ::| | : |  |
Db    409 GGLLDEAMLDEISL 422
```

RESULT 8
T10171
phospholipase D (EC 3.1.4.4) - castor bean
N;Alternate names: choline phosphatase; lecithinase D; lipophosphodiesterase II
C;Species: Ricinus communis (castor bean)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 13-Aug-1999
C;Accession: T10171; T10118; A54850
R;Xu, L.; Zheng, L.; Coughlan, S.J.; Wang, X.
Plant Mol. Biol. 32, 767-771, 1996
A;Title: Structure and analysis of phospholipase D gene from Ricinus communis L.
A;Reference number: Z16972; MUID:97134969
A;Accession: T10171
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-808 <XUL>
A;Cross-references: EMBL:U72693; NID:g1698843; PID:g1698844
A;Experimental source: cultivar Hale; tissue_type leaf
A;Zheng, X.
Submitted to the EMBL Data Library, July 1996
A;Reference number: Z16957
A;Accession: T10118
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-267,'L',269-808 <WAN>
A;Cross-references: EMBL:L33686; NID:g1438074; PID:g1438075
A;Experimental source: strain Hale; tissue_type endosperm
R;Wang, X.; Xu, L.; Zheng, L.
J. Biol. Chem. 269, 20312-20317, 1994
A;Title: Cloning and expression of phosphatidylcholine-hydrolyzing phospholipase D from
A;Reference number: A54850; MUID:94327597
A;Accession: A54850
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-101,'P',103-267,'L',269-447,'HLLLC',453-456,'LRHGMSSCLDPLMVELHLVSLRHLKMRQR
A;Cross-references: GB:L33686
A;Note: the authors translated the codon AAG for residue 535 as Leu
C;Genetics:
A;Introns: 31/3; 665/1
C;Keywords: phosphoric diester hydrolase

```
  Query Match             6.1%;  Score 107;  DB 2;  Length 808;
  Best Local Similarity   24.6%;  Pred. No. 1;
  Matches   59;  Conservative   24;  Mismatches   81;  Indels   76;  Gaps   14;

Qy    120 VHAVFSAPAVPSGTGQTSAELEVQ---RRHSLVSFVVRIVPSPD-------WFVGVDSLD 169
          || |   |  |   |    :|::      |  :  |   :|: |        |||   ||
Db    303 VHCVL-CPRNPDDGGSFVQDLQISTMFTHHQKIVVVDSAMPNGDSQRRRIVSFVG--GLD 359

Qy    170 LCDGDRWREQAALDLYPYDAGTDSGF---------TFSSPNFATIPQDTVTEITSSSPSH 220
          ||||             ||:   | |          |  ||||         |    |
Db    360 LCDG------------RYDSPFHSLFRTLDSAHHDDFHQPNFAG------ASIEKGGPRE 401

Qy    221 PANSFYYPRLKALPPIA---------------RVTLVRLRQSPRAFIPPAPVLPSRDNE 264
          | :   : ||:   |||                :  |::||:     |||:||:   | |
Db    402 PWHDI-HSRLEG--PIAWDVLFNFEQRWRKQGGKDLLIQLRELEDVIIPPSPVMYPDDFE 458

Qy    265 I--------VDSASV---PETPLDCEVSLWSSWGLCGGHCGRLGTKSRTRYVR-VQPANN 312
              :|  :    |||| |   :     ||  |    :    :  |:  :: | |
Db    459 AWNVQLFRSIDGGAAFGFPETPEDAPEA-----GLVSGKDNIIDRSIQDAYIHAIRRAKN 513

RESULT 9
T24232
hypothetical protein R17.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T24232
R;Barlow, K.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z19860
A;Accession: T24232
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-440 <WIL>
A;Cross-references: EMBL:Z92809; PIDN:CAB07269.1; GSPDB:GN00021; CESP:R17.3
A;Experimental source: clone R17
C;Genetics:
A;Gene: CESP:R17.3
A;Map position: 3
A;Introns: 36/3; 89/3; 179/2; 241/1; 306/1; 346/2
C;Superfamily: Caenorhabditis elegans hypothetical protein R17.3

```
  Query Match   6.0%; Score 104.5; DB 2; Length 440;
  Best Local Similarity   25.0%; Pred. No. 0.8;
  Matches   44; Conservative   20; Mismatches   43; Indels   69; Gaps   11;

Qy   206 PQDTVT-EITSSSPSHPANSFYYPRLKAL----------PPIARVTLVRLRQSPR----- 249
          |:| | || ::|: |||| | || : | : ||
Db   131 PEDVETQEIPLTTPA--------PRLKTLIFGLKRDQRTPP---EEITRYSEKPRHLLIR 179

Qy   250 -----AFIP----PAPVL-------PSRDNEIVDSASVPE------------------- 273
               |:| |:| |: :   ::| |: |
Db   180 YSLLSKFMPLKVTSTPLLYEENRVQPANNLYYLES-SISECYCDEHCVTLGDCCSDYTFV 238

Qy   274 -TPLDCEVSLWSSWGLC---GGHCGRLGTKSRTRYVRVQPANNGSPCPELEEEAEC 325
           | || :: | || | | || :||: | |:| |: | |:| |
Db   239 CPPRDCVLTDWDSWTQCTADNGTCG-IGTQKRLRHVIQHAERGGAACEPLKEMRTC 293

RESULT 10
A57293
latent transforming growth factor beta-binding protein 3 precursor - mouse
N;Alternate names: mitosis-inhibitory peptide
C;Species: Mus musculus (house mouse)
C;Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 11-Jan-2000
C;Accession: A57293; A60487
R;Yin, W.; Smiley, E.; Germiller, J.; Mecham, R.P.; Florer, J.B.; Wenstrup, R.J.; Bon
J. Biol. Chem. 270, 10147-10160, 1995
A;Title: Isolation of a novel latent transforming growth factor-beta binding protein
A;Reference number: A57293; MUID:95247723
A;Accession: A57293
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1251 <YIN>
A;Cross-references: GB:L40459
R;Reichelt, K.L.; Paulsen, J.E.; Elgjo, K.
Virchows Arch. B Cell Pathol. 59, 137-142, 1990
A;Title: Isolation of a growth and mitosis inhibitory peptide from mouse liver.
A;Reference number: A60487
A;Accession: A60487
A;Molecule type: protein
A;Residues: 65-69 <REI>
C;Comment: The molecular source of this pentapeptide has not been shown but it corres
C;Genetics:
A;Gene: Ltbp-3
C;Superfamily: unassigned EGF-related proteins; EGF homology
C;Keywords: liver; pyroglutamic acid
F;338-373/Domain: EGF homology <EGF>
F;65/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experi

```
  Query Match             5.9%;  Score 104;  DB 2;  Length 1251;
  Best Local Similarity   24.3%;  Pred. No. 3.1;
  Matches   76;  Conservative   23;  Mismatches  100;  Indels  114;  Gaps   18;

Qy     30 GGESICSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQWSSLLGAAHSSDYSMWRKNQY 89
          ||:  ||:|        |||:: |          ||  :   ||   |    |
Db    100 GGQ--CSSRNQCLCPPDFTGRFCQV----------PAAGT---GAGTGSSGPGWPDRAM 143

Qy     90 VSNGLRDFAERGEAWALMKEIEAA---------GEALQSVHAVFSAPAVPSGTGQTSAEL 140
```

```
Db    144 STGPLPPLAPEGESVASKHAIYAVQVIADPPGPGEGPPAQHAAF---LVPLGPGQISA-- 198

Qy    141 EVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQAALDLYPYDAGTDSGFTFSSP 200
          |||    :|:  ||:   |:  | |  ::                              |
Db    199 EVQAPPPVVN--VRVHHPPEASVQVHRIE----------------------------GP 227

Qy    201 NFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVTLVRLRQSP--RAFIPPAPVL 258
          | |  |  :   :    | ||            ||          | |  | |    |
Db    228 N-AEGPASSQHLLPHPKPQHPR----------PP---------TQKPLGRCFQDTLPKQ 266

Qy    259 PSRDNEIVDSASVPETPL-------DCEVSLWSSWGLCGGH-CGRLGTKSRTRYVRVQ-- 308
          |   |           ||       ||  |: ::||    | | :|      :|  ||
Db    267 PCGSN----------PLPGLTKQEDCCGSIGTAWGQSKCHKCPQL------QYTGVQKP 309

Qy    309 -PANN--GSPCPE 318
           |    |: ||:
Db    310 VPVRGEVGADCPQ 322
```

RESULT 11
BPI3
minor coat protein - phage Pf3
C;Species: phage Pf3
A;Note: host Pseudomonas aeruginosa
C;Date: 28-May-1986 #sequence_revision 28-May-1986 #text_change 23-Jul-1999
C;Accession: A04232
R;Luiten, R.G.M.; Putterman, D.G.; Schoenmakers, J.G.G.; Konings, R.N.H.; Day, L.A.
J. Virol. 56, 268-276, 1985
A;Title: Nucleotide sequence of the genome of Pf3, an IncP-1 plasmid-specific filamentou
A;Reference number: A94693; MUID:85293231
A;Accession: A04232
A;Molecule type: DNA
A;Residues: 1-483 <LUI>
A;Cross-references: GB:M11912; NID:g215371; PIDN:AAA88379.1; PID:g215374
C;Comment: Bacteriophage Pf3 is a class II filamentous phage.
C;Comment: The host is strain O harboring IncP1 plasmids.
C;Genetics:
A;Gene: 483
C;Superfamily: class II filamentous phage minor coat protein
C;Keywords: capsid protein

  Query Match 5.8%; Score 102; DB 1; Length 483;
  Best Local Similarity 25.8%; Pred. No. 1.4;
  Matches 62; Conservative 24; Mismatches 84; Indels 70; Gaps 14;

```
Qy      7 AAALGKALCALLLAT------LGAAGQPL--GGESICSARALAKYSITFTGKWSQTAFPK 58
          |: :|  | ||| |       |   :||  ||  :   | :     | ||   | :  |
Db    112 ASGIGLGLAALLEAADWVFDEEGEIVKPLPGGGSPVLMPRPVILNEYTVTGSAGQWSISK 171

Qy     59 QY-PLFRPPAQWSSLLG------AAHSSDYSMWR---------KNQYVSNGLRDFAERG- 101
          :| |  |    | |  |       |     :: ||           |   | |  ::: |
Db    172 EYEPDPRSVPGWYSYNGNPVWVSAVEDVGFT-WRYWYFADVLMDGQGRPNYLVAYSDSGP 230

Qy    102 -EAWALMKEIEAAGEALQSVHAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPD 160
           | |      :  | :| |:      |  |     | ||||
Db    231 NEYWQ-----DVGGYSLDSL------PTEPEFVPLTDAELE------------------ 260

Qy    161 WFVGVDSLDLCDGDRWREQAALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSH 220
            |:|     | | ||      :|:||    || ||  :|    ||   :: : ||| ::
Db    261 --AGIDQYYEPDPDDWR-----NLFPY-IEPDS-FTIETP----IPSLDLSPVVSSSTNN 307
```

RESULT 12
D96566
hypothetical protein F6D8.21 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: D96566
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: D96566
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-810 <STO>
A;Cross-references: GB:AE005173; NID:g5903048; PIDN:AAD55607.1; GSPDB:GN00141
C;Genetics:
A;Gene: F6D8.21
A;Map position: 1

  Query Match 5.7%; Score 100.5; DB 2; Length 810;
  Best Local Similarity 25.7%; Pred. No. 3.5;
  Matches 47; Conservative 17; Mismatches 40; Indels 79; Gaps 10;

```
Qy    129 VPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQAALDLYPYD 188
          :|||           |   :|||             |  ||||||             ||
Db    342 MPSGG---------SRSRRIVSF-------------VGGLDLCDG-----------RYD 367

Qy    189 AGTDSGF---------TFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIA-- 237
              | |          |  |||      |   ||   |  | :  :  ||:   |||
Db    368 TPFHSLFRTLDTAHHDDFHQPNF------TGAAITKGGPREPWHDIHC-RLEG--PIAWD 418

Qy    238 --------------RVTLVRLRQSPRAFIPPAPVLPSRDNEI--------VD---SASVP 272
                       :  ||::|:     |||:||| | |:::          :|   :|  |
Db    419 VLYNFEQRWSRQGGKDILVKMRELGDIIIPPSPVLFSEDHDVWNVQLFRSIDGGAAAGFP 478

Qy    273 ETP 275
          ::|
Db    479 DSP 481
```

RESULT 13
T33152
hypothetical protein T04D1.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33152
R;Davidson, S.; Wohldmann, P.
submitted to the EMBL Data Library, May 1998
A;Description: The sequence of C. elegans cosmid T04D1.
A;Reference number: Z21292
A;Accession: T33152
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2957 <DAV>
A;Cross-references: EMBL:AF067617; PIDN:AAC17559.1; GSPDB:GN00019; CESP:T04D1.4
A;Experimental source: strain Bristol N2; clone T04D1
C;Genetics:
A;Gene: CESP:T04D1.4
A;Map position: 1
A;Introns: 122/3; 293/3; 515/3; 1205/2; 1577/1; 2221/3; 2776/1; 2864/3

  Query Match 5.7%; Score 100.5; DB 2; Length 2957;
  Best Local Similarity 18.8%; Pred. No. 17;
  Matches 63; Conservative 55; Mismatches 154; Indels 63; Gaps 10;

```
Qy      7 AAALGKALCALLLATLGAAGQPLGGES-------ICSARALAKYSITFTGKWSQTAFPKQ 59
          | || :   |   :|  |  | :  :          |     |  ::| |   : |: |
Db   2407 AVALAQVAAAASASTSSADAQAVAQQQQVLEALLAMSMNPAALAALTGTSSSTGTSTPSA 2466

Qy     60 YPLFRPPAQWSSLLGA--------AHSSDYSMWRKNQYVSNGLRDFAERG------EAWA 105
            :|  |  |   |         | ::  :  :: |  :   :  |::       |
Db   2467 PKKSKPAPQKESTSSATAAAAAASAQAAALAAAQQQQQQAQQAQQAAQQAQSAKTQEELL 2526
```

```
Qy     106 LMKEIEAAGEALQ---SVHAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWF 162
           ::: :| ||  :|    :: :     ||    ||   | | :|  :    | :: :|:|
Db    2527 ILRLLELAGVGMQELAKLNTMSKDARVPMIHKQTRESLPVSKRPPIRDLTVFVMSNPEWS 2586

Qy     163 VGVDSLDLCDGDRWREQAALDLYPYDAGTDSGFTFSSPNFATIPQDTVT--EITSSSPSH 220
           : :   :   |:               || ||   ||   :: |    |   : :  ||
Db    2587 IDIPLFNEMSGN--------------GTGSGAAPSSKAASSRPTPAATPLAVATPKPST 2631

Qy     221 P--ANSFYYPRLKALPPIARVTLVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDC 278
           |   |   |     |  :    ::: :| : |     |   :  |:: ::  |:
Db    2632 PKVEKSSATPSAPETPSASATCNIKIGESLQ-FDDMISVFNRKTGELLAASKWPKA---T 2687

Qy     279 EVSLW---------------SSWGLCGGHCGRLG 297
           |:| |               :   | | |  |:|
Db    2688 ELSAWLDANPDYNVHANSALFAHLSLGGAHSDRIG 2722


RESULT  14
C70651
hypothetical protein Rv3074 - Mycobacterium tuberculosis (strain H37RV)
Species: Mycobacterium tuberculosis
Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: C70651
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: C70651
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-424 <COL>
A;Cross-references: GB:Z83866; GB:AL123456; NID:g3261691; PIDN:CAB06249.1; PID:e290921;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3074

  Query Match             5.7%;  Score 99.5;  DB 2;  Length 424;
  Best Local Similarity   23.4%;  Pred. No. 1.9;
  Matches   93;  Conservative   29;  Mismatches  136;  Indels  139;  Gaps   22;

Qy       1 MENPSPAAALGKALCALLL-----ATLGAAGQPLG--GESICSARALAKYSITFTGKWSQ 53
           :|    ||| |:|  |  :     |  |||| |    |  : |  |||:
Db      24 LERLKSAAAAGQARAAAAVDAARRAAEGAAGVPAARRGRGLASEIALARRD--------- 74

Qy      54 TAFPKQYPLFRPPAQWSSLLGAAHSSDYSM-------------WRKNQYV-SNGLRDFA 98
                   ||: |  || | :  | |              ||    |  :   | |
Db      75 -----------SPARGSRHLGFAKALVYEMPHTLAALDCGALSEWRATLIVRESACLDVA 123

Qy      99 ER----GEAWALMKEIEAAGEALQSVHAVFSAPAV-----PSGTGQTSAELEVQRRHSLV 149
           :|      |      ::|  |:|      | :| |:     |      :|  |  |
Db     124 DRRALDAELCGDPGDLEGMGDA----RVVAAARAIAYRLDPQAVVDRAANAENDR----- 174

Qy     150 SFVVRIVPSPDWFVGVDSL-----------------DLCDGDRWREQAALD-LYPYDAG 190
             | |:||    : :|                   ||| | | |   | |     |
Db     175 --TVTIRPAPDTMTYLTALLPVAQGVSVYAALTRAADTRCDG-RSRGQVMADTLVERVTG 231

Qy     191 TDSGFTFSSPNFATIPQDTVTEITSS------SPSHPANSFYY---PRLKALPPIA---- 237
           |          | :|      :  |      : : ||    |   |   |   :|
Db     232 RD----------AAVPTPIAVNLVMSDETLLGAANTPAQLCGYGPIPAAVARTMVASAVT 281

Qy     238 ----RVTLVRLRQSP----------RAFIPP---APVLPSRDNEIVDSASVPETPLDCEV 280
               | || ||   |           || : |   |  :  ||           ||  |:
Db     282 DQRSRATLRRLYAHPQAGALVSMESRARLFPRGLAAFIELRDQRC-------RTPY-CDA 333

Qy     281 SL--------WSSWGLCGGHCGRLGTKSRTRYVRVQP 309
           :        |:  |    | | |||  |  | :  |
Db     334 PIRHRDHAHPWADGGPTSAHNG-LGTCERCNYAKQAP 369


RESULT  15
I46687
complement component C8 beta subunit - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 20-Aug-1999
C;Accession: I46687
R;White, R.V.; Kaufman, K.M.; Letson, C.S.; Platteborze, P.L.; Sodetz, J.M.
J. Immunol. 152, 2501-2508, 1994
A;Title: Characterization of rabbit complement component C8: Functional evidence for
A;Reference number: I46686; MUID:94179833
A;Accession: I46687
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-590 <WHI>
A;Cross-references: GB:L26980; NID:g469062; PIDN:AAA31192.1; PID:g469063
C;Superfamily: complement C9; EGF homology; LDL receptor ligand-binding repeat homolo
F;63-117/Domain: thrombospondin type 1 repeat homology <THR>
F;122-155/Domain: LDL receptor ligand-binding repeat homology <LDL>
F;503-534/Domain: EGF homology <EGF>

  Query Match             5.7%;  Score 99.5;  DB 2;  Length 590;
  Best Local Similarity   33.3%;  Pred. No. 2.9;
  Matches   28;  Conservative   15;  Mismatches   28;  Indels   13;  Gaps    6;

Qy     246 QSPRAFIPPA---PVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKSRT 302
           : ||:  |      :  || :  ||:  :   |:|||:| ||||  :|   |      | |
Db      33 ERPRSLEPTVVNRSLAKSRHSRSVDATPM---PIDCELSSWSSWTMC-DPC----QKKRY 84

Qy     303 RYV-RVQPAN-NGSPCPELEEEAE 324
           |:   ::|:  || ||   ::| |
Db      85 RHAYLLRPSQFNGEPCNFSDKEVE 108


Search completed: December 10, 2001, 20:03:09
Job time: 2731 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2001, 19:46:05 ; Search time 62.4 Seconds
(without alignments)
194.488 Million cell updates/sec

Title: US-09-170-042A-2
Perfect score: 1756
Sequence: 1 MENPSPAAALGKALCALLLA.....NGSPCPELEEEAECVPDNCV 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

[redacted]tal number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

*Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	462.5	26.3	807	1	FSPO_RAT	P35446 rattus norv
2	448.5	25.5	803	1	FSPO_XENLA	P35447 xenopus lae
3	127.5	7.3	741	1	NFL1_MOUSE	Q61985 mus musculu
4	117	6.7	772	1	NFL1_HUMAN	Q14494 h nuclear f
5	107	6.1	808	1	PLD_RICCO	Q41142 ricinus com
6	102	5.8	483	1	COA2_BPPF3	P03624 bacteriopha
7	100.5	5.7	810	1	PLD2_ARATH	Q9ssq9 arabidopsis
8	99.5	5.7	590	1	CO8B_RABIT	P98137 oryctolagus
9	99.5	5.7	812	1	PLD_MAIZE	Q43270 zea mays (m
10	98.5	5.6	812	1	PLD2_BRAOC	P55939 brassica ol
11	97.5	5.6	808	1	PLD_TOBAC	P93400 nicotiana t
12	97	5.5	1584	1	BAI1_HUMAN	O14514 homo sapien
13	96.5	5.5	598	1	BGAL_XANMN	P48982 xanthomonas
14	96.5	5.5	810	1	PLD1_BRAOC	O82549 brassica ol
15	96.5	5.5	812	1	PLD1_ORYSA	Q43007 oryza sativ
16	96	5.5	591	1	CO8B_HUMAN	P07358 homo sapien
17	94	5.4	809	1	PLD_VIGUN	O04865 vigna ungui
18	93.5	5.3	810	1	PLD1_ARATH	Q38882 arabidopsis
19	93	5.3	3027	1	POLG_PYFV1	Q05057 parsnip yel
20	92.5	5.3	912	1	PGCB_BOVIN	Q28062 bos taurus
21	92	5.2	1306	1	MSB2_YEAST	P32334 saccharomyc
22	91.5	5.2	341	1	GUN1_CRYFL	Q04469 cryptococcu
23	90	5.1	2265	1	FINC_BOVIN	P07589 bos taurus
24	89	5.1	1257	1	PER2_MOUSE	O54943 mus musculu
25	89	5.1	1567	1	FMN2_MOUSE	Q9j104 mus musculu
26	88	5.0	867	1	SSPO_BOVIN	P98167 bos taurus
27	87.5	5.0	897	1	CHIA_ECOLI	P13656 escherichia
28	87	5.0	283	1	ISPE_CHLMU	Q9plc0 chlamydia m
29	87	5.0	541	1	BBR1_SCHCO	P78741 schizophyll
30	87	5.0	863	1	AMPN_CAUCR	P37893 caulobacter
31	86.5	4.9	400	1	TX22_HUMAN	Q9y458 homo sapien
32	86.5	4.9	590	1	INR1_MOUSE	P33896 mus musculu
33	86	4.9	940	1	MAZ4_SCHCO	P37938 schizophyll
34	85.5	4.9	1572	1	BAI2_HUMAN	Q60241 homo sapien
35	85	4.8	1093	1	SM5B_MOUSE	Q60519 mus musculu
36	85	4.8	1094	1	EX5B_MYCTU	P96920 mycobacteri
37	85	4.8	1114	1	RET_HUMAN	P07949 homo sapien
38	83.5	4.8	1675	1	CLH1_HUMAN	Q00610 homo sapien
39	83.5	4.8	1675	1	CLH_BOVIN	P49951 bos taurus
40	83.5	4.8	1675	1	CLH_RAT	P11442 rattus norv
41	83.5	4.8	2333	1	PGCA_CANFA	Q28343 canis famil
42	83	4.7	888	1	UFO_MOUSE	Q00993 mus musculu
43	82.5	4.7	279	1	GLPF_PSEAE	Q51389 pseudomonas
44	82.5	4.7	437	1	ERR1_YEAST	P42222 saccharomyc
45	82.5	4.7	1395	1	IF4G_HUMAN	Q04637 homo sapien

ALIGNMENTS

```
RESULT   1
FSPO_RAT
ID   FSPO_RAT        STANDARD;      PRT;   807 AA.
AC   P35446;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   F-SPONDIN PRECURSOR.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryonic floor plate;
RX   MEDLINE=92208952; PubMed=1555244;
RA   Klar A., Baldassare M., Jessell T.M.;
RT   "F-spondin: a gene expressed at high levels in the floor plate
RT   encodes a secreted protein that promotes neural cell adhesion and
RT   neurite extension.";
RL   Cell 69:95-110(1992).
CC   -!- FUNCTION: PROMOTES THE ATTACHMENT OF SPINAL CORD AND SENSORY
CC       NEURON CELLS AND THE OUTGROWTH OF NEURITES IN VITRO. MAY
CC       CONTRIBUTE TO THE GROWTH AND GUIDANCE OF AXONS IN BOTH THE SPINAL
CC       CORD AND THE PNS.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THE FLOOR PLATE.
CC   -!- SIMILARITY: CONTAINS 6 TSP TYPE-1 DOMAINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M88469; AAA41174.1; -.
DR   PIR; A38152; A38152.
DR   InterPro; IPR002861; Reeler.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF02014; Reeler; 1.
DR   Pfam; PF00090; tsp_1; 6.
DR   SMART; SM00209; TSP1; 6.
DR   PROSITE; PS50092; TSP1; 5.
KW   Glycoprotein; Signal; Repeat; Cell adhesion.
FT   SIGNAL       1     28       POTENTIAL.
FT   CHAIN       29    807       F-SPONDIN.
FT   DOMAIN     443    494       TSP TYPE-1 1.
FT   DOMAIN     502    554       TSP TYPE-1 2.
FT   DOMAIN     559    610       TSP TYPE-1 3.
FT   DOMAIN     615    665       TSP TYPE-1 4.
FT   DOMAIN     669    720       TSP TYPE-1 5.
FT   DOMAIN     755    807       TSP TYPE-1 6.
FT   CARBOHYD   214    214       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   681    681       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
SQ   SEQUENCE   807 AA;  90773 MW;  309525F9EAFEA89A CRC64;

  Query Match             26.3%;  Score 462.5;  DB 1;  Length 807;
  Best Local Similarity   34.2%;  Pred. No. 4e-30;
  Matches  116;  Conservative   49;  Mismatches  137;  Indels   37;  Gaps   12;

Qy       9 ALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQ 68
           :| | ||       |   :|:     | |   ||| :|| | ||:   || ||  |
Db     176 SLTKKLCEQDPTLDGVTDRPI---LDCCACGTAKYRLTFYGNWSEKTHPKDYP--RRANH 230

Qy      69 WSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEI-EAAGEALQSVHAVFSAP 127
           ||:::| :|| :| :|    | | |::  || |    : :|| : : | |  : |    |
Db     231 WSAIIGGSHSKNYVLWEYGGYASEGVKQVAELGSPVKMEEEIRQQSDEVLTVIKAKAQWP 290

Qy     128 --------AVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGD-RWRE 178
                   | |      |||  | |   |:||:  : |||| ||: : |||  :  | :
Db     291 SWQPVNVRAAP------SAEFSVDRTRHLMSFLTMMGPSPDWNVGLSAEDLCTKECGWVQ 344

Qy     179 QAALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIAR 238
           :   || |:|||||| |: ||| ||||: :  :||    || : || |   ::  :||
Db     345 KVVQDLIPWDAGTDSGVTYESPNKPTIPQEKIRPLTSL--DHPQSPFYDPEGGSITQVAR 402

Qy     239 VTLVRL-RQSPRAFIPPAPVLPSRDNEIVDSASVPE------TPLDCEVSLWSSWGLCGG 291
           | : |: |:  :  | |  |    |: : | |  ||      ||  |  | || |  |
Db     403 VVIERIARKGEQCNIVPDNV----DDIVADLA--PEEKDEDDTPETCIYSNWSPWSACSS 456

Qy     292 HCGRLGTKSRTRYVRVQPANNGSPCPELEEEAECVPDNC 330
             | : | | :: |  :   |||: ::   |:   |
Db     457 STCEKGKRMRQRMLKAQ-LDLSVPCPDTQDFQPCMGPGC 494


RESULT   2
FSPO_XENLA
ID   FSPO_XENLA     STANDARD;      PRT;   803 AA.
AC   P35447;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   F-SPONDIN PRECURSOR.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=93376785; PubMed=8367492;
RA   Ruiz I Altaba A., Cox C., Jessell T.M., Klar A.;
RT   "Ectopic neural expression of a floor plate marker in frog embryos
RT   injected with the midline transcription factor Pintallavis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:8268-8272(1993).
CC   -!- FUNCTION: PROMOTES THE ATTACHMENT OF SPINAL CORD AND SENSORY
CC       NEURON CELLS AND THE OUTGROWTH OF NEURITES IN VITRO. MAY
CC       CONTRIBUTE TO THE GROWTH AND GUIDANCE OF AXONS IN BOTH THE SPINAL
CC       CORD AND THE PNS.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THE FLOOR PLATE.
CC   -!- SIMILARITY: CONTAINS 6 TSP TYPE-1 DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L09123; AAA19105.1; -.
DR   PIR; A47723; A47723.
DR   InterPro; IPR002861; Reeler.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF02014; Reeler; 1.
DR   Pfam; PF00090; tsp_1; 6.
DR   SMART; SM00209; TSP1; 6.
DR   PROSITE; PS50092; TSP1; 6.
KW   Glycoprotein; Signal; Repeat; Cell adhesion.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    803       F-SPONDIN.
FT   DOMAIN      437    488       TSP TYPE-1 1.
FT   DOMAIN      496    548       TSP TYPE-1 2.
FT   DOMAIN      553    604       TSP TYPE-1 3.
FT   DOMAIN      609    661       TSP TYPE-1 4.
FT   DOMAIN      665    716       TSP TYPE-1 5.
FT   DOMAIN      751    803       TSP TYPE-1 6.
FT   CARBOHYD    210    210       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    677    677       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   803 AA;  90702 MW;  D3A54E329548AED9 CRC64;

  Query Match             25.5%;  Score 448.5;  DB 1;  Length 803;
  Best Local Similarity   34.3%;  Pred. No. 5.4e-29;
  Matches  115;  Conservative   48;  Mismatches  141;  Indels   31;  Gaps   12;

Qy       9 ALGKALCALLLATLGAAGQPLGGESI---CSARALAKYSITFTGKWSQTAFPKQYPLFRP 65
           :| | :| | |   |      | |     | |   ||| :|| | ||: | || ||  |
Db     172 SLTKRMCELDLTLEG------GNEKTIPDCCACGTAKYRLTFYGNWSEKAHPKDYP--RR 223

Qy      66 PAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAG-EALQSVHAVF 124
             ||:::| :|| :| :|   |  |:|::  || |    : :||   | | |  : |
Db     224 ANHWSAIIGGSHSGEYVLWEYGQ-ASDGVKQVAELGSPVKMEEEIRQKGDEVLTVIKAKA 282

Qy     125 SAPAVP--SGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGD-RWREQAA 181
             ||    :     |||  | |   |:||:  : ||||| ||: | |||  :  | ::
Db     283 QWPAWQPLNVRAAPSAEFSVDRSRHLMSFLAMMGPSPDWNVGLTSEDLCTKECGWVQKVV 342

Qy     182 LDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITS-SSPSHPANSFYYPRLKALPPIARVT 240
           || |:|||||| |: |||  |||| :  :||   |  |: :    |   : |||||
Db     343 QDLIPWDAGTDSGVTYESPNKPTIPQDKIRPLTSLDHPQSPSMT----RGGPIIPIARVV 398

Qy     241 LVRLRQSPRAFIPPAPVLPSRDNEIV-----DSASVPETPLDCEVSLWSSWGLCGGHCGR 295
           : |: :          ::|   ::||     :     :||  |  | || |  |
Db     399 IERIARKGE----QCNIIPDNVDDIVADLVTEEKDEDDTPETCIYSNWSPWSACSSATCD 454

Qy     296 LGTKSRTRYVRVQPANNGSPCPELEEEAECVPDNC 330
           | : | | :: |  :   |||: ::   |:   |
Db     455 KGKRMRQRMLKAQ-LDLSVPCPDTQDFEPCMGPGC 488


RESULT   3
NFL1_MOUSE
ID   NFL1_MOUSE     STANDARD;      PRT;   741 AA.
AC   Q61985; O70234;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   NUCLEAR FACTOR ERYTHROID 2 RELATED FACTOR 1 (NF-E2 RELATED FACTOR 1)
DE   (NFE2-RELATED FACTOR 1) (NUCLEAR FACTOR, ERYTHROID DERIVED 2, LIKE 1).
GN   NFE2L1 OR NRF1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=95278942; PubMed=7759107;
RA   McKie J., Johnstone K., Mattei M.-G., Scambler P.;
RT   "Cloning and mapping of murine Nfe2l1.";
RL   Genomics 25:716-719(1995).
RN   [2]
RP   SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RX   MEDLINE=98248571; PubMed=9580677;
RA   Prieschl E.E., Novotny V., Csonga R., Jaksche D., Elbe-Buerger A.,
RA   Thumb W., Auer M., Stingl G., Baumruker T.;
```

```
RT   "A novel splice variant of the transcription factor Nrf1 interacts
RT   with the TNFalpha promoter and stimulates transcription.";
RL   Nucleic Acids Res. 26:2291-2297(1998).
CC   -!- FUNCTION: THE SHORT ISOFORM INTERACTS WITH THE EXTENDED KAPPA 3
CC       SITE OF THE TNF ALPHA PROMOTER AFTER FC GAMMA RIII STIMULATION AND
CC       PARTICIPATES IN THE INDUCTION OF THIS CYTOKINE. THE LONG ISOFORM
CC       IS EITHER INACTIVE OR REPRESSES THE TRANSCRIPTIONAL ACTIVATION.
CC   -!- SUBUNIT: HETERODIMER (PROBABLE). IT MAY FORM A HETERODIMER WITH
CC       SEVERAL TRANSCRIPTION FACTORS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG (SHOWN HERE) AND A
CC       SHORT; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC   -!- PTM: COULD BE DEPENDENT ON CKII PHOSPHORYLATION FOR BINDING.
CC   -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. CNC SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X78709; CAA55362.1; -.
DR   EMBL; AF015881; AAC40108.1; -.
DR   HSSP; P34707; 1SKN.
DR   MGD; MGI:99421; Nfe211.
DR   InterPro; IPR001871; bZIP.
DR   Pfam; PF00170; bZIP; 1.
DR   SMART; SM00338; BRLZ; 1.
DR   PROSITE; PS00036; BZIP_BASIC; 1.
KW   DNA-binding; Nuclear protein; Phosphorylation; Alternative splicing.
FT   DOMAIN      125    259       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      414    447       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      467    486       POLY-SER.
FT   DNA_BIND    628    643       BASIC MOTIF (BY SIMILARITY).
FT   DOMAIN      651    673       LEUCINE-ZIPPER (BY SIMILARITY).
FT   DOMAIN      730    737       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   VARSPLIC      1    291       MISSING (IN SHORT ISOFORM).
FT   VARSPLIC    447    583       MISSING (IN SHORT ISOFORM).
FT   CONFLICT    318    318       S -> T (IN REF. 2).
FT   CONFLICT    387    387       P -> L (IN REF. 2).
SQ   SEQUENCE   741 AA;  81545 MW;  C01E89DD26E7CDEF CRC64;

  Query Match             7.3%;  Score 127.5;  DB 1;  Length 741;
  Best Local Similarity   22.4%;  Pred. No. 0.0073;
  Matches   79;  Conservative   45;  Mismatches  119;  Indels  109;  Gaps   17;

Qy      19 LATLGAAGQPLGGESICSARALAKYSITFTG-KWSQTAFPKQYPLFRPPAQWSSLLGAAH 77
           |  |||   |: |:       | |  |      | |                   |||
Db     153 LEDLGAVAPPVSGD-------LTKEDIDLIDILWRQDID---------------LGAGR 189

Qy      78 S-SDYSMWRKNQYVSNGLRDFAERGEAW------ALMKEIEAAGEALQSVHAVFSA--PA 128
             |||  :| | |   |:|  || : |       || :::   ||  :|  | | |   :
Db     190 EVFDYSHRQKEQDVDKELQDGREREDTWSGEGAEALARDLLVDGETGESFPAQFPADVSS 249

Qy     129 VPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWRE-QAALDLYPY 187
           :|      |   :|  :||:| ::    ||         ||    :|::  : :::
Db     250 IPEAVPSESESPALQ--NSLLSPLLTGTESP--------FDL--EQQWQDLMSIMEMQAM 297

Qy     188 DAGTD-SGFTFSSPNFATIPQDTVTEITSSSPSHPAN----------------SFYYPR 229
           :  |  |   :::|     | | ::   | :|: | |                 | : |
Db     298 EVNTSASEILYNAP-----PGDPLSSNYSLAPNTPINQNVSLHQASLGGCSQDFSLFSPE 352

Qy     230 LKALPPIARVTLVRLRQSPRAFIP--------PAPVLPSRDNEIVDSASVPETP------ 275
           :::||  :  ||: | |    :           |  ||: |   :  | || |
Db     353 VESLPVASSSTLLPLVPSNSTSLNSTFGSTNLAGPFFPSQLNGTANDTSGPELPDPLGGL 412

Qy     276 -----LDCEVSLWSSWGLCGGHCGRLGTKSRTRYVRVQPANNGSPCPELEEE 322
               || |:||                    : ::   |     :||||
Db     413 LDEAMLD-EISLMD-------------------LAIEEGFNPVQASQLEEE 443
```

```
RESULT   4
NFL1_HUMAN
ID   NFL1_HUMAN     STANDARD;      PRT;   772 AA.
AC   Q14494; Q12877;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   NUCLEAR FACTOR ERYTHROID 2 RELATED FACTOR 1 (NF-E2 RELATED FACTOR 1)
DE   (NFE2-RELATED FACTOR 1) (NUCLEAR FACTOR, ERYTHROID DERIVED 2, LIKE 1)
DE   (TRANSCRIPTION FACTOR 11) (TRANSCRIPTION FACTOR HBZ17) (TRANSCRIPTION
DE   FACTOR LCR-F1) (LOCUS CONTROL REGION-FACTOR 1).
GN   NFE2L1 OR NRF1 OR TCF11 OR HBZ17.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95095252; PubMed=8001966;
RA   Luna L., Johnsen O., Skartlien A.H., Pedeutour F., Turc-Carel C.,
RA   Prydz H., Kolstoe A.-B.;
RT   "Molecular cloning of a putative novel human bZIP transcription
RT   factor on chromosome 17q22.";
RL   Genomics 22:553-562(1994).
RN   [2]
RP   SEQUENCE OF 326-772 FROM N.A.
RX   MEDLINE=94310069; PubMed=8036168;
RA   Caterina J.J., Donze D., Sun C.W., Ciavatta D.J., Townes T.M.;
RT   "Cloning and functional characterization of LCR-F1: a bZIP
RT   transcription factor that activates erythroid-specific, human globin
RT   gene expression.";
RL   Nucleic Acids Res. 22:2383-2391(1994).
CC   -!- FUNCTION: ACTIVATES ERYTHROID-SPECIFIC, GLOBIN GENE EXPRESSION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. CNC SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X77366; CAA54555.1; -.
DR   EMBL; U08853; AAA20466.1; -.
DR   HSSP; P34707; 1SKN.
DR   MIM; 163260; -.
DR   InterPro; IPR001871; bZIP.
DR   Pfam; PF00170; bZIP; 1.
DR   SMART; SM00338; BRLZ; 1.
DR   PROSITE; PS00036; BZIP_BASIC; 1.
KW   Transcription regulation; Activator; DNA-binding; Nuclear protein.
FT   DOMAIN      125    288       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      496    517       POLY-SER.
FT   DNA_BIND    659    674       BASIC MOTIF.
FT   DOMAIN      682    704       LEUCINE-ZIPPER.
SQ   SEQUENCE   772 AA;  84703 MW;  C868807C6046BEF5 CRC64;

  Query Match             6.7%;  Score 117;  DB 1;  Length 772;
  Best Local Similarity   23.3%;  Pred. No. 0.055;
  Matches   78;  Conservative   35;  Mismatches  110;  Indels  112;  Gaps   17;

Qy      19 LATLGAAGQPLGGESICSARALAKYSITFTG-KWSQTAFPKQYPLFRPPAQWSSLLGAAH 77
           |  |||   |: |:       | |  |      | |                   |||
Db     153 LEDLGAVAPPVSGD-------LTKEDIDLIDILWRQDID---------------LGAGR 189

Qy      78 S-SDYSMWRKNQYVSNGLRDFAER----GE-AWALMKEIEAAGEALQSVHAVFSAPAVPS 131
             |||  :| | |   |||  |:    || | || : :   ||  :|  |      |||
Db     190 EVFDYSHRQKEQDVEKELRDGGEQDTWAGEGAEALARNLLVDGETGESFPA-----QVPS 244
```

```
Qy     132 GTGQTSAELEVQRR--------------HSLVSFVVRIVPSPD------------WFVGV 165
           |  ||:  ||   |               : :| :   |||                 |
Db     245 GEDQTALSLEECLRLLEATCPFGENAEFPADISSITEAVPSESEPPALQNNLLSPLLTGT 304

Qy     166 DS-LDLCDGDRWRE-QAALDLYPYDAGTD-SGFTFSSPNFATIPQDTVTEITSSSPSHPA 222
           :|  ||    :|::  : :::   :  |  |   :|:|     | | ::   | :|: |
Db     305 ESPFDL--EQQWQDLMSIMEMQAMEVNTSASEILYSAP-----PGDPLSTNYSLAPNTPI 357

Qy     223 NS-----------------FYYPRLKALPPIARVTLVRLRQSPRA--------------F 251
           |                 : | :::||  :  ||: |  |                  |
Db     358 NQNVSLHQASLGGCSQDFLLFSPEVESLPVASSSTLLPLAPSNSTSLNSTFGSTNLTGLF 417

Qy     252 IPP----------APVLPSRDNEIVDSASVPETPL 276
           ||           | ||     ::| | : |  |
Db     418 FPPQLNGTANDTAGPELPDPLGGLLDEAMLDEISL 452


RESULT   5
PLD_RICCO
ID   PLD_RICCO      STANDARD;      PRT;   808 AA.
AC   Q41142; P93507;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   PHOSPHOLIPASE D PRECURSOR (EC 3.1.4.4) (PLD) (CHOLINE PHOSPHATASE)
DE   (PHOSPHATIDYLCHOLINE-HYDROLYZING PHOSPHOLIPASE D).
OS   Ricinus communis (Castor bean).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Malpighiales; Euphorbiaceae; Ricinus.
OX   NCBI_TaxID=3988;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=CV. HALE; TISSUE=Endosperm;
RX   MEDLINE=94327597; PubMed=8051126;
RA   Wang X., Xu L., Zheng L.;
RT   "Cloning and expression of phosphatidylcholine-hydrolyzing
RT   phospholipase D from Ricinus communis L.";
RL   J. Biol. Chem. 269:20312-20317(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Leaf;
RX   MEDLINE=97134969; PubMed=8980529;
RA   Xu L., Zheng L., Coughlan S.J., Wang X.;
RT   "Structure and analysis of phospholipase D gene from Ricinus communis
RT   L.";
RL   Plant Mol. Biol. 32:767-771(1996).
CC   -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN CELLULAR PATHWAYS INCLUDING
CC       SIGNAL TRANSDUCTION PATHWAYS.
CC   -!- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O = CHOLINE + A
CC       PHOSPHATIDATE.
CC   -!- COFACTOR: CALCIUM (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: MOSTLY EXPRESSED IN VACUOLES, ENDOPLASMIC
CC       RETICULUM, A FEW IN PLASTIDS AND PLASMA MEMBRANE. EXPRESSION IS
CC       HIGHER IN RADICLE THAN IN ENDOSPERM.
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L33686; AAB04095.1; -.
DR   EMBL; U72693; AAB37305.1; -.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001736; PLD.
DR   Pfam; PF00614; PLDc; 2.
DR   SMART; SM00239; C2; 1.
DR   SMART; SM00155; PLDc; 2.
DR   PROSITE; PS50004; C2_DOMAIN_2; 1.
KW   Hydrolase; Lipid degradation; Calcium.
FT   PROPEP        1     30
FT   CHAIN        31    808       PHOSPHOLIPASE D.
FT   DOMAIN        1    109       C2 DOMAIN.
FT   CONFLICT    268    268       L -> I (IN REF. 2).
SQ   SEQUENCE   808 AA;  91992 MW;  E75F6CFFB9ADF3CB CRC64;


  Query Match             6.1%;  Score 107;  DB 1;  Length 808;
  Best Local Similarity   24.6%;  Pred. No. 0.38;
  Matches   59;  Conservative   24;  Mismatches   81;  Indels   76;  Gaps   14;

Qy     120 VHAVFSAPAVPSGTGQTSAELEVQ---RRHSLVSFVVRIVPSPD-------WFVGVDSLD 169
           || |   |  |   |    :|::      |  :  |   :|: |         |||   ||
Db     303 VHCVL-CPRNPDDGGSFVQDLQISTMFTHHQKIVVVDSAMPNGDSQRRRIVSFVG--GLD 359

Qy     170 LCDGDRWREQAALDLYPYDAGTDSGF---------TFSSPNFATIPQDTVTEITSSSPSH 220
           ||||              ||:   | |          |  ||||          |    |
Db     360 LCDG------------RYDSPFHSLFRTLDSAHHDDFHQPNFAG------ASIEKGGPRE 401

Qy     221 PANSFYYPRLKALPPIA----------------RVTLVRLRQSPRAFIPPAPVLPSRDNE 264
           | :   : ||:   |||                :  |::||:     |||:||:   | |
Db     402 PWHDI-HSRLEG--PIAWDVLFNFEQRWRKQGGKDLLIQLRELEDVIIPPSPVMYPDDFE 458

Qy     265 I--------VDSASV---PETPLDCEVSLWSSWGLCGGHCGRLGTKSRTRYVR-VQPANN 312
           :|  :    |||| |   :     ||  |    :    :  |:  :: | |
Db     459 AWNVQLFRSIDGGAAFGFPETPEDAPEA-----GLVSGKDNIIDRSIQDAYIHAIRRAKN 513


RESULT   6
COA2_BPPF3
ID   COA2_BPPF3     STANDARD;      PRT;   483 AA.
AC   P03624;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-MAR-1989 (Rel. 10, Last annotation update)
DE   MINOR COAT PROTEIN (ORF 483).
OS   Bacteriophage Pf3.
OC   Viruses; ssDNA viruses; Inoviridae; Inovirus.
OX   NCBI_TaxID=10872;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NEW-YORK, AND NIJMEGEN;
RX   MEDLINE=85293231; PubMed=3928901;
RA   Luiten R.G.M., Putterman D.G., Schoenmakers J.G.G., Konings R.N.H.,
RA   Day L.A.;
RT   "Nucleotide sequence of the genome of Pf3, an IncP-1 plasmid-specific
RT   filamentous bacteriophage of Pseudomonas aeruginosa.";
RL   J. Virol. 56:268-276(1985).
CC   -!- MISCELLANEOUS: THE STRAIN NIJMEGEN SEQUENCE IS SHOWN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M11912; AAA88379.1; -.
DR   EMBL; M19377; AAA88388.1; -.
DR   PIR; A04232; VCBPI3.
KW   Coat protein.
FT   VARIANT     419    419       L -> F (IN STRAIN NEW-YORK).
SQ   SEQUENCE   483 AA;  52197 MW;  DCBBAA598BA700BB CRC64;


  Query Match             5.8%;  Score 102;  DB 1;  Length 483;
  Best Local Similarity   25.8%;  Pred. No. 0.52;
  Matches   62;  Conservative   24;  Mismatches   84;  Indels   70;  Gaps   14;

Qy       7 AAALGKALCALLLAT------LGAAGQPL--GGESICSARALAKYSITFTGKWSQTAFPK 58
```

```
          |: :|  | |||  |        |   :||  ||  :   | :     | ||   | :  |
Db    112 ASGIGLGLAALLEAADWVFDEEGEIVKPLPGGGSPVLMPRPVILNEYTVTGSAGQWSISK 171

Qy     59 QY-PLFRPPAQWSSLLG------AAHSSDYSMWR---------KNQYVSNGLRDFAERG- 101
          :| |  |    | |  |      |     :: ||           |   | |  ::: |
Db    172 EYEPDPRSVPGWYSYNGNPVWVSAVEDVGFT-WRYWYFADVLMDGQGRPNYLVAYSDSGP 230

Qy    102 -EAWALMKEIEAAGEALQSVHAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPD 160
           | |      :  | :| |:      |  |     | ||||
Db    231 NEYWQ-----DVGGYSLDSL------PTEPEFVPLTDAELE------------------- 260

Qy    161 WFVGVDSLDLCDGDRWREQAALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSH 220
            |:|     | | ||     :|:||    || ||  :|    ||   :: : ||| ::
Db    261 --AGIDQYYEPDPDDWR-----NLFPY-IEPDS-FTIETP----IPSLDLSPVVSSSTNN 307


RESULT   7
PLD2_ARATH
ID   PLD2_ARATH     STANDARD;      PRT;   810 AA.
AC   Q9SSQ9;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PHOSPHOLIPASE D 2 PRECURSOR (EC 3.1.4.4) (PLD 2) (CHOLINE PHOSPHATASE
DE   2) (PHOSPHATIDYLCHOLINE-HYDROLYZING PHOSPHOLIPASE D 2).
GN   PLD2 OR AT1G52570 OR F6D8.21.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
CC   -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN VARIOUS CELLULAR PROCESSES.
CC   -!- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O = CHOLINE + A
CC       PHOSPHATIDATE.
CC   -!- COFACTOR: CALCIUM (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AC008016; AAD55607.1; -.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001736; PLD.
DR   Pfam; PF00614; PLDc; 2.
DR   SMART; SM00239; C2; 1.
DR   SMART; SM00155; PLDc; 2.
DR   PROSITE; PS50004; C2_DOMAIN_2; FALSE_NEG.
KW   Hydrolase; Lipid degradation; Calcium.
FT   PROPEP          1     36       BY SIMILARITY.
FT   CHAIN          37    810       PHOSPHOLIPASE D 2.
FT   DOMAIN          1    110       C2 DOMAIN.
SQ   SEQUENCE   810 AA;  91597 MW;  F1FA36D8C9CFBCCA CRC64;


  Query Match             5.7%;  Score 100.5;  DB 1;  Length 810;
  Best Local Similarity   25.7%;  Pred. No. 1.3;
  Matches   47;  Conservative   17;  Mismatches   40;  Indels   79;  Gaps   10;

Qy    129 VPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQAALDLYPYD 188
          :|||           |   :|||             |  ||||||             ||
Db    342 MPSGG---------SRSRRIVSF------------VGGLDLCDG-----------RYD 367

Qy    189 AGTDSGF---------TFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIA-- 237
              | |         |  |||      |   ||   |  | :  :  ||:   |||
Db    368 TPFHSLFRTLDTAHHDDFHQPNF------TGAAITKGGPREPWHDIHC-RLEG--PIAWD 418

Qy    238 --------------RVTLVRLRQSPRAFIPPAPVLPSRDNEI--------VD---SASVP 272
                       :  ||::|:     |||:||| | |:::         :|   :|  |
Db    419 VLYNFEQRWSRQGGKDILVKMRELGDIIIPPSPVLFSEDHDVWNVQLFRSIDGGAAAGFP 478

Qy    273 ETP 275
          ::|
Db    479 DSP 481


RESULT   8
CO8B_RABIT
ID   CO8B_RABIT     STANDARD;      PRT;   590 AA.
AC   P98137;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COMPLEMENT COMPONENT C8 BETA CHAIN PRECURSOR.
GN   C8B.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NEW ZEALAND WHITE; TISSUE=Liver;
RX   MEDLINE=94179833; PubMed=7510745;
RA   White R.V., Kaufman K.M., Letson C.S., Platteborze P.L.,
RA   Sodetz J.M.;
RT   "Characterization of rabbit complement component C8. Functional
RT   evidence for the species-selective recognition of C8 alpha by
RT   homologous restriction factor (CD59).";
RL   J. Immunol. 152:2501-2508(1994).
CC   -!- FUNCTION: C8 IS A CONSTITUENT OF THE MEMBRANE ATTACK COMPLEX.
CC       C8 BINDS TO THE C5B-7 COMPLEX, FORMING THE C5B-8 COMPLEX. C5-B8
CC       BINDS C9 AND ACTS AS A CATALYST IN THE POLYMERIZATION OF C9.
CC   -!- SUBUNIT: C8 IS COMPOSED OF THREE CHAINS: ALPHA, BETA AND GAMMA.
CC       THE BETA CHAIN BINDS TO THE C8 ALPHA CHAIN AND TO THE C5B-C7
CC       COMPLEX, PRESUMABLY TO C5B. IT IS ESSENTIAL TO THE INCORPORATION
CC       OF C8 INTO THE C5B-C8 COMPLEX.
CC   -!- SIMILARITY: TO COMPLEMENT FACTORS C6, C7, C9, AND TO PERFORIN.
CC   -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; L26980; AAA31192.1; -.
DR   HSSP; P00750; 1TPG.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR002172; LDL_recept_A.
DR   InterPro; IPR001862; MAC_perforin.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00057; ldl_recept_a; 1.
DR   Pfam; PF01823; MACPF; 1.
DR   Pfam; PF00090; tsp_1; 2.
DR   PRINTS; PR00764; COMPLEMENTC9.
DR   SMART; SM00192; LDLa; 1.
DR   SMART; SM00457; MACPF; 1.
DR   SMART; SM00209; TSP1; 2.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01209; LDLRA_1; 1.
DR   PROSITE; PS50068; LDLRA_2; 1.
DR   PROSITE; PS00279; MAC_PERFORIN; FALSE_NEG.
DR   PROSITE; PS50092; TSP1; 2.
KW   Complement pathway; Complement alternate pathway; Glycoprotein;
     Signal; Plasma; Membrane attack complex; Cytolysis; EGF-like domain;
     Repeat; Transmembrane.
FT   SIGNAL        1     26       POTENTIAL.
FT   PROPEP       27     54       POTENTIAL.
FT   CHAIN        55    590       COMPLEMENT COMPONENT C8 BETA CHAIN.
FT   DOMAIN       64    119       TSP TYPE-1 1.
FT   DOMAIN      120    157       LDL-RECEPTOR CLASS A.
FT   TRANSMEM    312    328       POTENTIAL.
FT   TRANSMEM    333    352       POTENTIAL.
FT   DOMAIN      499    535       EGF-LIKE.
FT   DOMAIN      542    590       TSP TYPE-1 2.
FT   DISULFID    122    133       BY SIMILARITY.
FT   DISULFID    127    146       BY SIMILARITY.
FT   DISULFID    140    155       BY SIMILARITY.
FT   DISULFID    378    403       BY SIMILARITY.
FT   CARBOHYD     44     44       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    101    101       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   590 AA;  66857 MW;  54ED81B15F2CCC2C CRC64;


  Query Match             5.7%;  Score 99.5;  DB 1;  Length 590;
  Best Local Similarity   33.3%;  Pred. No. 1.1;
  Matches   28;  Conservative   15;  Mismatches   28;  Indels   13;  Gaps    6;

Qy     246 QSPRAFIPPA---PVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKSRT 302
           : ||:  |      :  || :  ||:  :   |:|||:| |||| :|   |      | |
Db      33 ERPRSLEPTVVNRSLAKSRHSRSVDATPM---PIDCELSSWSSWTMC-DPC----QKKRY 84

       303 RYV-RVQPAN-NGSPCPELEEEAE 324
           |:   ::|:  || ||   ::| |
Db      85 RHAYLLRPSQFNGEPCNFSDKEVE 108


RESULT   9
PLD_MAIZE
ID   PLD_MAIZE      STANDARD;      PRT;   812 AA.
AC   Q43270;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   PHOSPHOLIPASE D PRECURSOR (EC 3.1.4.4) (PLD) (CHOLINE PHOSPHATASE)
DE   (PHOSPHATIDYLCHOLINE-HYDROLYZING PHOSPHOLIPASE D).
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. MO17;
RX   MEDLINE=96012933; PubMed=7551587;
RA   Ueki J., Morioka S., Komari T., Kumashiro T.;
RT   "Purification and characterization of phospholipase D (PLD) from rice
RT   (Oryza sativa L.) and cloning of cDNA for PLD from rice and maize
RT   (Zea mays L.).";
RL   Plant Cell Physiol. 36:903-914(1995).
CC   -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN VARIOUS CELLULAR PROCESSES.
CC   -!- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O = CHOLINE + A
CC       PHOSPHATIDATE.
CC   -!- COFACTOR: CALCIUM (BY SIMILARITY).
CC   -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D73410; BAA11135.1; -.
DR   MaizeDB; 113853; -.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001736; PLD.
DR   Pfam; PF00614; PLDc; 2.
DR   SMART; SM00239; C2; 1.
DR   SMART; SM00155; PLDc; 2.
DR   PROSITE; PS50004; C2_DOMAIN_2; 1.
KW   Hydrolase; Lipid degradation; Calcium.
FT   PROPEP       1      ?       POTENTIAL.
FT   CHAIN        ?    812       PHOSPHOLIPASE D.
FT   DOMAIN       1    114       C2 DOMAIN.
SQ   SEQUENCE   812 AA;  92242 MW;  D05CB351655BCC61 CRC64;


  Query Match             5.7%;  Score 99.5;  DB 1;  Length 812;
  Best Local Similarity   25.7%;  Pred. No. 1.6;
  Matches   49;  Conservative   19;  Mismatches   62;  Indels   61;  Gaps   12;

Qy     127 PAVPSGTGQTSAELEVQ---RRHSLVSFVVRIVPSPD-------WFVGVDSLDLCDGDRW 176
           |  |  :|    :|::      |  :  |   :|:          |:|   :|||||  |:
Db     313 PRNPDDSGSFVQDLQISTMFTHHQKIVVVDHEMPNQGSQQRRIVSFIG--GIDLCDG-RY 369

Qy     177 REQ-----AALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLK 231
             |      ||   :|       |  |||          |    |  | :   : ||:
Db     370 DTQYHSLFRTLDTVHHD-------DFHQPNFEG------GSIKKGGPREPWHDI-HSRLE 415

Qy     232 ALPPIA---------------RVTLVRLRQSPRAFIPPAPVLPSRDNEI--------VD 267
             |||                :  |||||  |   |||:||:   | |          :|
Db     416 G--PIAWDVLYNFEQRWRKQGGKDLLVRLRDLPDIIIPPSPVMFPEDRETWNVQLFRSID 473

Qy     268 SASV---PETP 275
           :    ||||
Db     474 GGAAFGFPETP 484


RESULT  10
PLD2_BRAOC
ID   PLD2_BRAOC     STANDARD;      PRT;   812 AA.
AC   P55939; O49981; O82548;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PHOSPHOLIPASE D 2 PRECURSOR (EC 3.1.4.4) (PLD 2) (CHOLINE PHOSPHATASE
DE   2) (PHOSPHATIDYLCHOLINE-HYDROLYZING PHOSPHOLIPASE D 2).
GN   PLD2.
OS   Brassica oleracea var. capitata (Cabbage).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI_TaxID=3716;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kim D.U., Lee J.E., Hoe K.R., Jang Y.J., Rho T.Y., Yoo H.S.,
RA   Choi M.-U.;
```

```
RT   "Cloning of phospholipase D from cabbage.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Pannenberg I., Mansfeld J., Ulbrich-Hofmann R.;
RT   "Cloning of phospholipase D from Brassica oleracea var. capitata.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 38-68.
RC   TISSUE=Leaf;
RX   MEDLINE=93357261; PubMed=8353126;
RA   Abousalham A., Riviere M., Teissere M., Verger R.;
RT   "Improved purification and biochemical characterization of
RT   phospholipase D from cabbage.";
RL   Biochim. Biophys. Acta 1158:1-7(1993).
CC   -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN VARIOUS CELLULAR PROCESSES.
CC   -!- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O = CHOLINE + A
CC       PHOSPHATIDATE.
CC   -!- COFACTOR: CALCIUM. OPTIMAL PH DEPENDS ON CALCIUM CONCENTRATION.
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U85482; AAC79125.1; -.
DR   EMBL; AF090444; AAC78486.1; -.
DR   EMBL; AF113919; AAD17209.1; -.
DR   Mendel; 27805; Braol;1491;27805.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001736; PLD.
DR   Pfam; PF00614; PLDc; 2.
DR   SMART; SM00239; C2; 1.
DR   SMART; SM00155; PLDc; 2.
DR   PROSITE; PS50004; C2_DOMAIN_2; FALSE_NEG.
KW   Hydrolase; Lipid degradation; Calcium.
FT   PROPEP        1     37
FT   CHAIN        38    812       PHOSPHOLIPASE D 2.
FT   DOMAIN        1    111       C2 DOMAIN.
FT   CONFLICT     50     50       Q -> E (IN REF. 3).
SQ   SEQUENCE   812 AA;  92061 MW;  F771FC53734E7A2B CRC64;
```

```
Query Match             5.6%;  Score 98.5;  DB 1;  Length 812;
Best Local Similarity   25.9%;  Pred. No. 1.9;
Matches   57;  Conservative   19;  Mismatches   67;  Indels   77;  Gaps   14;

Qy     120 VHAVFSAPAVPSGTGQTSAELEVQ---RRHSLVSFVVRIVPSPDW---------FVGVDS 167
           || |   |  |   |     |:|      |  :  |   |||             |||
Db     305 VHCVL-CPRNPDDGGSIVQNLQVSAMFTHHQKIVVVDSEVPSQGGGSEMRRIMSFVG--G 361

Qy     168 LDLCDGDRWREQAALDLYPYDAGTDSGF---------TFSSPNFATIPQDTVTEITSSSP 218
           :|||||             ||    | |          |  |||       |   ||   |
Db     362 IDLCDG------------RYDTPFHSLFRTLDTVHHDDFHQPNF------TGASITKGGP 403

Qy     219 SHPANSFYYPRLKALPPIA---------------RVTLVRLRQSPRAFIPPAPVLPSRD 262
             |     : ||:   |||                :  ||:||:     | |:||:   |
Db     404 REPWQDI-HSRLEG--PIAWDVLYNFEQRWSKQGGKDILVKLRELSDIIITPSPVMFQED 460

Qy     263 NEI--------VD---SASVPETPLDCEVSLWSSWGLCGG 291
           :::        :|   :|  |::|   ||:  :  ||  |
Db     461 HDVWNVQLFRSIDGGAAAGFPDSP---EVA--AEAGLVSG 495
```

```
RESULT  11
PLD_TOBAC
ID   PLD_TOBAC      STANDARD;      PRT;   808 AA.
AC   P93400; P93399;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   PHOSPHOLIPASE D PRECURSOR (EC 3.1.4.4) (PLD) (CHOLINE PHOSPHATASE)
DE   (PHOSPHATIDYLCHOLINE-HYDROLYZING PHOSPHOLIPASE D).
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. SR1; TISSUE=Leaf;
RA   Lein W., Saalbach G.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN CELLULAR PATHWAYS INCLUDING
CC       SIGNAL TRANSDUCTION PATHWAYS.
CC   -!- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O = CHOLINE + A
CC       PHOSPHATIDATE.
CC   -!- COFACTOR: CALCIUM (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z84822; CAB06620.1; -.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001736; PLD.
DR   Pfam; PF00168; C2; 1.
DR   Pfam; PF00614; PLDc; 2.
DR   SMART; SM00239; C2; 1.
DR   SMART; SM00155; PLDc; 2.
DR   PROSITE; PS50004; C2_DOMAIN_2; 1.
KW   Hydrolase; Lipid degradation; Calcium.
FT   PROPEP        1     30       BY SIMILARITY.
FT   CHAIN        31    808       PHOSPHOLIPASE D.
FT   DOMAIN        1    109       C2 DOMAIN.
SQ   SEQUENCE   808 AA;  91966 MW;  2937FCA5D5C60B6A CRC64;
```

```
Query Match             5.6%;  Score 97.5;  DB 1;  Length 808;
Best Local Similarity   24.0%;  Pred. No. 2.3;
Matches   56;  Conservative   23;  Mismatches   79;  Indels   75;  Gaps   13;

Qy     127 PAVPSGTGQTSAELEVQ---RRHSLVSFVVRIVPSPD-------WFVGVDSLDLCDGDRW 176
           |  |   |     |::     |  :  |   :|| :         |||   :|||||
Db     309 PRNPDDGGSIVQSLQIGTMFTHHQKIVVVDSELPSGESEKRRILSFVG--GIDLCDG--- 363

Qy     177 REQAALDLYPYDAGTDSGF---------TFSSPNFATIPQDTVTEITSSSPSHPANSFYY 227
                 ||    | |          |  |||   |   :|:     | |  :|
Db     364 ---------RYDTPFHSLFRTLDTAHHDDFHQPNF---PDGAITKGGPREPWHDIHS--- 408

Qy     228 PRLKALPPIA---------------RVTLVRLRQSPRAFIPPAPVLPSRDNEI------ 265
            ||:   |||                :  ||  |:     |||:||:   |:|
Db     409 -RLEG--PIAWDVLFNFEQRWRKQGGKDVLVNFRELDDIIIPPSPVMHLDDSETWNVQLF 465

Qy     266 --VDSASV---PETPLDCEVSLWSSWGLCGGHCGRLGTKSRTRYVR-VQPANN 312
             :|  :    |||| |      :  ||  |    :    :  |:  :: | |
Db     466 RSIDEGAAFGFPETPEDA-----AKAGLVSGXDNIIDRSIQDAYIHAIRRAKN 513
```

```
RESULT  12
BAI1_HUMAN
ID   BAI1_HUMAN      STANDARD;      PRT;  1584 AA.
AC   O14514;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   BRAIN-SPECIFIC ANGIOGENESIS INHIBITOR 1 PRECURSOR.
```

```
GN   BAI1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=98054121; PubMed=9393972;
RA   Nishimori H., Shiratsuchi T., Urano T., Kimura Y., Kiyono K.,
RA   Tatsumi K., Yoshida S., Ono M., Kuwano M., Nakamura Y., Tokino T.;
RT   "A novel brain-specific p53-target gene, BAI1, containing
RT   thrombospondin type 1 repeats inhibits experimental angiogenesis.";
RL   Oncogene 15:2145-2150(1997).
RN   [2]
RP   INTERACTION WITH BAP1.
RX   MEDLINE=98321173; PubMed=9647739;
RA   Shiratsuchi T., Futamura M., Oda K., Nishimori H., Nakamura Y.,
RA   Tokino T.;
RT   "Cloning and characterization of BAI-associated protein 1: a PDZ
RT   domain-containing protein that interacts with BAI1.";
RL   Biochem. Biophys. Res. Commun. 247:597-604(1998).
CC   -!- FUNCTION: LIKELY TO BE A POTENT INHIBITOR OF ANGIOGENESIS IN
CC       BRAIN AND MAY PLAY A SIGNIFICANT ROLE AS A MEDIATOR OF THE P53
CC       SIGNAL IN SUPPRESSION OF GLIOBLASTOMA. MAY FUNCTION IN CELL
CC       ADHESION AND SIGNAL TRANSDUCTION IN THE BRAIN.
CC   -!- SUBUNIT: INTERACTS WITH BAP1.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. LIKELY TO BE
CC       CONCENTRATED AT CELL-CELL ADHESION SITES.
CC   -!- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN BRAIN. REDUCED OR NO
CC       EXPRESSION IS OBSERVED IN SOME GLIOBLASTOMA CELL LINES AND CANCER
CC       TISSUES.
CC   -!- INDUCTION: BY P53.
CC   -!- DOMAIN: THE TSP1 REPEATS INHIBIT IN VIVO ANGIOGENESIS IN RAT
CC       CORNEA INDUCED BY BFGF.
CC   -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC   -!- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB005297; BAA23647.1; -.
DR   MIM; 602682; -.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR001879; hormn_rcptor.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF01825; GPS; 1.
DR   Pfam; PF00090; tsp_1; 5.
DR   SMART; SM00303; GPS; 1.
DR   SMART; SM00008; HormR; 1.
DR   SMART; SM00209; TSP1; 5.
DR   PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR   PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW   Repeat; Cell adhesion.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1584       BRAIN-SPECIFIC ANGIOGENESIS INHIBITOR 1.
FT   DOMAIN       31    948       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    949    969       1 (POTENTIAL).
FT   DOMAIN      970    980       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    981   1001       2 (POTENTIAL).
FT   DOMAIN     1002   1008       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1009   1029       3 (POTENTIAL).
FT   DOMAIN     1030   1052       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1053   1073       4 (POTENTIAL).
FT   DOMAIN     1074   1093       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1094   1114       5 (POTENTIAL).
FT   DOMAIN     1115   1136       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1137   1157       6 (POTENTIAL).
FT   DOMAIN     1158   1166       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1167   1187       7 (POTENTIAL).
FT   DOMAIN     1188   1584       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      261    316       TSP TYPE-1 1.
FT   DOMAIN      354    408       TSP TYPE-1 2.
FT   DOMAIN      409    463       TSP TYPE-1 3.
FT   DOMAIN      467    521       TSP TYPE-1 4.
FT   DOMAIN      522    576       TSP TYPE-1 5.
FT   DOMAIN     1411   1422       POLY-PRO.
FT   DOMAIN     1425   1430       POLY-PRO.
FT   SITE        231    233       CELL ATTACHMENT SITE (POTENTIAL).
FT   DOMAIN     1365   1584       NECESSARY FOR INTERACTION WITH BAP1.
FT   DOMAIN     1581   1584       INDISPENSABLE FOR INTERACTION WITH BAP1.
FT   CARBOHYD     64     64       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    401    401       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    607    607       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    692    692       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    844    844       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    877    877       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    881    881       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1584 AA;  173531 MW;  DEA8F28C77874513 CRC64;


 Query Match             5.5%;  Score 97;  DB 1;  Length 1584;
 Best Local Similarity  21.9%;  Pred. No. 5.6;
 Matches   85;  Conservative   30;  Mismatches  131;  Indels  142;  Gaps   20;

Qy       4 PSPAAALGKALCALLLATLG-----AAGQPLG-GESICSARALAKYSITFTGKWSQTAPP 57
           | |   |   |  |||  ||      |||   | |   |:     |    | | :|  |
Db       8 PGPVWILAPLL--LLLLLLGRRARAAAGADAGPGPEPCATLVQGK----FFGYFSAAAV- 60

Qy      58 KQYPLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEAL 117
             :|       |:  |       |::: |
Db      61 --FPANASRCSWT--LRNPDPRRYTLYMK------------------------------ 85

Qy     118 QSVHAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLD----LCDG 173
             |  ||   || |:          | :   ||    : |   ::||:| |    |||
Db      86 -----VAKAPVPCSGPGRV-------RTYQFDSF----LESTRTYLGVESFDEVLRLCDP 129

Qy     174 D---------------RWREQAALDLYPYDAGTDSGFTFSSPNFATI----PQDTVTE-- 212
                           | ::    |     ||        |  : :    |     :
Db     130 SAPLAFLQASKQFLQMRRQQPPQHDGLRPRAGPPGPTDDFSVEYLVVGNRNPSRAACQML 189

Qy     213 -------ITSSSPSHPANSFYYPRLKALPPIARVTLVRLRQSPRAFI------PPA-PVL 258
                 :  |  |||                      : |:| | :      | | |:
Db     190 CRWLDACLAGSRSSHPCG-------------------IMQTPCACLGGEAGGPAAGPLA 229

Qy     259 PSRDNEIVDSASVPETPLDCEVS---------------LWSSWGLCGGHCGRLGTKSRTR 303
           |  |  : |  :|   | :|  |               ||| || |   ||  | ::|||
Db     230 PRGDVCLRD--AVAGGPENCLTSLTQDRGGHGATGGWKLWSLWGECTRDCGG-GLQTRTR 286

Qy     304 YVRVQPANNGSPCP-ELEEEAECVPDNC 330
             |   |  |||  :|  : |
Db     287 TCLPAPGVEGGGCEGVLEEGRQCNREAC 314


RESULT  13
BGAL_XANMN
ID   BGAL_XANMN     STANDARD;      PRT;   598 AA.
AC   P48982;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   BETA-GALACTOSIDASE PRECURSOR (EC 3.2.1.23) (LACTASE).
GN   BGA.
OS   Xanthomonas manihotis.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xanthomonas.
OX   NCBI_TaxID=43353;
RN   [1]
```

```
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 22-38.
RC   STRAIN=7AS1;
RX   MEDLINE=96121416; PubMed=8563148;
RA   Taron C.H., Benner J.S., Hornstra L.J., Guthrie E.P.;
RT   "A novel beta-galactosidase gene isolated from the bacterium
RT   Xanthomonas manihotis exhibits strong homology to several eukaryotic
RT   beta-galactosidases.";
RL   Glycobiology 5:603-610(1995).
CC   -!- FUNCTION: PREFERENTIALLY HYDROLYZES BETA(1->3) GALACTOSYL LINKAGES
CC       OVER BETA(1->4) LINKAGES. HAS A PH OPTIMUM OF 4.5.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC       GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC   -!- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L35444; AAC41485.1; -.
DR   InterPro; IPR001944; Glyco_hydro_35.
DR   Pfam; PF01301; Glyco_hydro_35; 1.
DR   PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
KW   Hydrolase; Glycosidase; Signal.
FT   SIGNAL        1     21
FT   CHAIN        22    598       BETA-GALACTOSIDASE.
FT   ACT_SITE    184    184       PROTON DONOR (POTENTIAL).
FT   ACT_SITE    260    260       NUCLEOPHILE (POTENTIAL).
SQ   SEQUENCE   598 AA;  66085 MW;  DB4C3F05E01435BF CRC64;

  Query Match             5.5%;  Score 96.5;  DB 1;  Length 598;
  Best Local Similarity   26.8%;  Pred. No. 1.9;
  Matches   38;  Conservative   14;  Mismatches   53;  Indels   37;  Gaps    7;

Qy     161 WF-------VGVDSLDLCDGDRW--REQAALDLYPYDAGTDSGFTFSSPNFATIPQDTVT 211
           ||           |:    :   |  |:  : :|| :  ||  ||  :  ||   | |
Db     265 WFDHWGKPHAATDARQQAEEFEWILRQGHSANLYMFIGGTSFGF-MNGANFQNNPSDHYA 323

Qy     212 EITSS--------SPSHPANSFYYPRLKALPPIARVTLVRLRQSPRAFIPPAPVLPSRDN 263
             |:|          ||   |   |      ||||| |:          |||   |
Db     324 PQTTSYDYDAILDEAGHPTPKFALMR----DAIARVTGVQ---------PPALPAP---- 366

Qy     264 EIVDSASVPETPLDCEVSLWSS 285
           : : ::| |||    ||| :
Db     367 --ITTTTLPATPLRESASLWDN 386

RESULT  14
PLD1_BRAOC
ID   PLD1_BRAOC     STANDARD;      PRT;   810 AA.
AC   O82549;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PHOSPHOLIPASE D 1 PRECURSOR (EC 3.1.4.4) (PLD 1) (CHOLINE PHOSPHATASE
DE   1) (PHOSPHATIDYLCHOLINE-HYDROLYZING PHOSPHOLIPASE D 1).
GN   PLD1.
OS   Brassica oleracea var. capitata (Cabbage).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI_TaxID=3716;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Pannenberg I., Mansfeld J., Ulbrich-Hofmann R.;
RT   "Cloning of phospholipase D from Brassica oleracea var. capitata.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN VARIOUS CELLULAR PROCESSES.
CC   -!- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O = CHOLINE + A
CC       PHOSPHATIDATE.
CC   -!- COFACTOR: CALCIUM (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF090445; AAC78487.1; -.
DR   EMBL; AF113918; AAD17208.1; -.
DR   Mendel; 33136; Braol;1491;33136.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001736; PLD.
DR   Pfam; PF00614; PLDc; 2.
DR   SMART; SM00239; C2; 1.
DR   SMART; SM00155; PLDc; 2.
DR   PROSITE; PS50004; C2_DOMAIN_2; FALSE_NEG.
KW   Hydrolase; Lipid degradation; Calcium.
FT   PROPEP        1     36       BY SIMILARITY.
FT   CHAIN        37    810       PHOSPHOLIPASE D 1.
FT   DOMAIN        1    110       C2 DOMAIN.
SQ   SEQUENCE   810 AA;  91837 MW;  CBC5B3D7E1F0DAAB CRC64;

  Query Match             5.5%;  Score 96.5;  DB 1;  Length 810;
  Best Local Similarity   25.1%;  Pred. No. 2.7;
  Matches   51;  Conservative   19;  Mismatches   62;  Indels   71;  Gaps   12;

Qy     120 VHAVFSAPAVPSGTGQTSAELEVQ---RRHSLVSFVVRIVPSPDW--------FVGVDSL 168
           || :   |  |   |     |:|       |  :  |   :||           |||   :
Db     304 VHCIL-CPRNPDDGGSIVQNLQVSAMFTHHQKIVVVDSEMPSRGGSQMRRIVSFVG--GI 360

Qy     169 DLCDGDRWREQAALDLYPYDAGTDSGF---------TFSSPNFATIPQDTVTEITSSSPS 219
           |||||            ||    | |         |  |||      |   ||   |
Db     361 DLCDG------------RYDTPFHSLFRTLDTVHHDDFHQPNF------TGAAITKGGPR 402

Qy     220 HPANSFYYPRLKALPPIA---------------RVTLVRLRQSPRAFIPPAPVLPSRDN 263
             |  :   : ||:   |||                :  ||:||:     | |:||:   |:
Db     403 EPWHDI-HSRLEG--PIAWDVLYNFEQRWSKQGGKDILVKLRELSDIIITPSPVMFQEDH 459

Qy     264 EI--------VD---SASVPETP 275
           ::        :|   :|  ||:|
Db     460 DVWNVQLFRSIDGGAAAGFPESP 482

RESULT  15
PLD1_ORYSA
ID   PLD1_ORYSA     STANDARD;      PRT;   812 AA.
AC   Q43007;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   PHOSPHOLIPASE D 1 PRECURSOR (EC 3.1.4.4) (PLD 1) (CHOLINE PHOSPHATASE
DE   1) (PHOSPHATIDYLCHOLINE-HYDROLYZING PHOSPHOLIPASE D 1).
GN   PLD1.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=CV. KOSHIHIKARI;
RX   MEDLINE=96012933; PubMed=7551587;
RA   Ueki J., Morioka S., Komari T., Kumashiro T.;
RT   "Purification and characterization of phospholipase D (PLD) from rice
RT   (Oryza sativa L.) and cloning of cDNA for PLD from rice and maize
RT   (Zea mays L.).";
RL   Plant Cell Physiol. 36:903-914(1995).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. KOSHIHIKARI; TISSUE=Leaf;
RA   Morioka S., Ueki J., Komari T.;
RT   "Characterization of two Distinctive genomic clones for phospholipase
RT   D from rice.";
RL   (In) Plant Gene Register PGR97-076.
CC   -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN VARIOUS CELLULAR PROCESSES.
CC       THERMOSTABLE BETWEEN 4 AND 37 DEGREES CELSIUS, BUT ACTIVITY IS
CC       REDUCED TO ONE THIRD OF THE ORIGINAL LEVEL AFTER INCUBATION AT 50
CC       DEGREES CELSIUS FOR 30 MINUTES. OPTIMAL PH IS 6.0.
CC   -!- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O = CHOLINE + A
CC       PHOSPHATIDATE.
CC   -!- COFACTOR: CALCIUM.
CC   -!- SUBUNIT: MONOMER.
CC   -!- TISSUE SPECIFICITY: EXPRESS IN LEAVES, ROOTS, DEVELOPING SEEDS AND
CC       CULTURED CELLS.
CC   -!- DEVELOPMENTAL STAGE: THE TRANSCRIPT BEGINS TO EMERGED IN SEEDS AS
CC       EARLY AS THE SECOND DAY AFTER IMBIBITION AND INCREASED AFTER
CC       RADICLE EMERGENCE ON THE THIRD DAY. STRONGLY EXPRESSED IN THE
CC       LEAVES, ROOTS AND RESIDUAL GRAIN OF SEEDLING EIGHT DAYS AFTER
CC       IMBIBITION. UNDETECTABLE IN MATURE DRY SEEDS.
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D73411; BAA11136.1; -.
DR   EMBL; AB001920; BAA19467.1; -.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001736; PLD.
DR   Pfam; PF00614; PLDc; 2.
DR   SMART; SM00239; C2; 1.
DR   SMART; SM00155; PLDc; 2.
DR   PROSITE; PS50004; C2_DOMAIN_2; 1.
KW   Hydrolase; Lipid degradation; Calcium; Multigene family.
FT   PROPEP          1     46
FT   CHAIN          47    812       PHOSPHOLIPASE D 1.
FT   DOMAIN          1    114       C2 DOMAIN.
SQ   SEQUENCE   812 AA;  92209 MW;  635F94E56CDEFD8F CRC64;


  Query Match             5.5%;  Score 96.5;  DB 1;  Length 812;
  Best Local Similarity   25.7%;  Pred. No. 2.7;
  Matches   49;  Conservative   19;  Mismatches   62;  Indels   61;  Gaps   12;

Qy     127 PAVPSGTGQTSAELEVQ---RRHSLVSFVVRIVPSPD-------WFVGVDSLDLCDGDRW 176
           |  |  :|    :| :      |  :  |   :|:          |||   |||||| |:
Db     313 PRNPDDSGSIVQDLSISTMFTHHQKIVVVDHELPNQGSQQRRIVSFVG--GLDLCDG-RY 369

Qy     177 REQ-----AALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLK 231
             |       ||   :|        |  |||||        |    |  | :   : ||:
Db     370 DTQYHSLFRTLDSTHHD-------DFHQPNFAT------ASIKKGGPREPWHDI-HSRLE 415

Qy     232 ALPPIA---------------RVTLVRLRQSPRAFIPPAPVLPSRDNEI--------VD 267
             |||               :  |::||      |||:||:   | |          :|
Db     416 G--PIAWDVLYNFEQRWRKQGGKDLLLQLRDLSDTIIPPSPVMFPEDRETWNVQLFRSID 473

Qy     268 SASV---PETP 275
           :    |:||
Db     474 GGAAFGFPDTP 484


Search completed: December 10, 2001, 20:04:23
Job time: 1098 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2001, 19:14:28 ; Search time 101.35 Seconds
(without alignments)
477.712 Million cell updates/sec

Title: US-09-170-042A-2
Perfect score: 1756
Sequence: 1 MENPSPAAALGKALCALLLA.....NGSPCPELEEEAECVPDNCV 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1749	99.6	331	4	Q9ULW1	Q9ulw1 homo sapien
2	1741	99.1	331	4	Q9BUD6	Q9bud6 homo sapien
3	1520.5	86.6	330	11	Q9WV75	Q9wv75 rattus norv
4	1108	63.1	331	13	O42112	O42112 brachydanio
5	868	49.4	334	13	O42111	O42111 brachydanio
6	528.5	30.1	601	5	Q9V746	Q9v746 drosophila
7	518	29.5	598	5	O02029	O02029 drosophila
8	470.5	26.8	808	13	O42113	O42113 brachydanio
9	464.5	26.5	802	13	Q9W770	Q9w770 gallus gall
10	460.5	26.2	807	4	Q9HCB6	Q9hcb6 homo sapien
11	460.5	26.2	807	6	Q9GLX9	Q9glx9 bos taurus
12	457	26.0	624	4	O94862	O94862 homo sapien
13	457	26.0	898	5	O76822	O76822 branchiosto
14	445.5	25.4	763	5	Q9XZD0	Q9xzd0 drosophila
15	443	25.2	803	13	O42114	O42114 brachydanio
16	439.5	25.0	873	5	Q9V692	Q9v692 drosophila
17	422.5	24.1	1682	5	Q9V693	Q9v693 drosophila
18	420.5	23.9	216	4	Q9H7I1	Q9h7i1 homo sapien
19	406	23.1	805	5	Q19305	Q19305 caenorhabdi
20	144.5	8.2	1536	4	Q9C0I4	Q9c0i4 homo sapien
21	135	7.7	446	4	O43384	O43384 homo sapien
22	134	7.6	1290	4	Q9UPZ6	Q9upz6 homo sapien
23	128	7.3	687	5	Q23729	Q23729 cryptospori
24	125	7.1	660	5	Q23832	Q23832 cryptospori
25	123.5	7.0	749	11	Q9Z2E8	Q9z2e8 mus musculu
26	121.5	6.9	238	5	O76510	O76510 cryptospori
27	117.5	6.7	770	5	Q20942	Q20942 caenorhabdi
28	114	6.5	1423	5	Q9W1A0	Q9w1a0 drosophila
29	112	6.4	543	4	O60407	O60407 homo sapien
30	109	6.2	168	5	Q9GZ22	Q9gz22 cryptospori
31	108	6.2	123	5	Q27550	Q27550 cryptospori
32	108	6.2	123	5	Q9TVV7	Q9tvv7 cryptospori
33	108	6.2	765	5	Q9U8J9	Q9u8j9 neospora ca
34	107.5	6.1	1523	3	Q9HFX4	Q9hfx4 candida alb
35	106	6.0	168	5	Q9GZ23	Q9gz23 cryptospori
36	105	6.0	123	5	O76782	O76782 cryptospori
37	105	6.0	123	5	Q9TZP9	Q9tzp9 cryptospori
38	104.5	6.0	440	5	O18003	O18003 caenorhabdi
39	103	5.9	265	11	Q9Z132	Q9z132 mus musculu
40	103	5.9	1253	11	Q61810	Q61810 mus musculu
41	102.5	5.8	588	13	Q9PVW7	Q9pvw7 paralichthy
42	102.5	5.8	738	5	Q9W3Z1	Q9w3z1 drosophila
43	102.5	5.8	809	10	Q9SDZ6	Q9sdz6 lycopersico
44	102	5.8	168	5	Q9GZ21	Q9gz21 cryptospori
45	102	5.8	1863	11	Q9JLQ1	Q9jlq1 mus musculu

ALIGNMENTS

RESULT 1
Q9ULW1
ID Q9ULW1 PRELIMINARY; PRT; 331 AA.
AC Q9ULW1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SPONDIN 2.
GN SPON2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99443867; PubMed=10512675;
RA Manda R., Kohno T., Matsuno Y., Takenoshita S., Kuwano H., Yokota J.;
RT "Identification of genes (SPON2 and C20orf2) differentially expressed
RT between cancerous and noncancerous lung cells by mRNA differential
RT display.";
RL Genomics 61:5-14(1999).
DR EMBL; AB027466; BAA85892.1; -.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp_1; 1.
DR PROSITE; PS50092; TSP1; 1.
DR SMART; SM00209; TSP1; 1.
SQ SEQUENCE 331 AA; 35772 MW; F1D6F0BAC1412CA7 CRC64;

Query Match 99.6%; Score 1749; DB 4; Length 331;
Best Local Similarity 99.7%; Pred. No. 6.4e-140;
Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQY 60
Db 1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARAPAKYSITFTGKWSQTAFPKQY 60

Qy 61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120
Db 61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120

Qy 121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180

Qy     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240

Qy     241 LVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 LVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300

Qy     301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
           |||||||||||||||||||||||||||||||
Db     301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331


RESULT   2
Q9BUD6
ID   Q9BUD6      PRELIMINARY;      PRT;   331 AA.
AC   Q9BUD6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SPONDIN 2, EXTRACELLULAR MATRIX PROTEIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CHORIOCARCINOMA;
RA   Strausberg R.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC002707; AAH02707.1; -.
KW   Matrix protein.
SQ   SEQUENCE   331 AA;  35844 MW;  418E244B893C59F4 CRC64;


  Query Match             99.1%;  Score 1741;  DB 4;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 3e-139;
  Matches  328;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQY 60
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1 MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARAPAKYSITFTGKWSQTAFPKQY 60

Qy      61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 PLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSV 120

Qy     121 HAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 HEVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA 180

Qy     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVT 240

Qy     241 LVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300
           |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 LLRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTKS 300

Qy     301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
           |||||||||||||||||||||||||||||||
Db     301 RTRYVRVQPANNGSPCPELEEEAECVPDNCV 331


RESULT   3
Q9WV75
ID   Q9WV75      PRELIMINARY;      PRT;   330 AA.
AC   Q9WV75;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MINDIN PRECURSOR.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE DAWLEY;
RA   Feinstein Y., Borrell V., Garcia C., Burstyn-Cohen T., Tzarfaty V.,
RA   Frumkin A., Nose A., Okamoto H., Higashijima S., Soriano A., Klar A.;
RT   "F-spondin and mindin: two structurally and functionally related genes
RT   expressed in the hippocampus that promote outgrowth of embryonic
RT   hippocampal neurons.";
RL   Development 0:0-0(1999).
DR   EMBL; AF155196; AAD38195.1; -.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00090; tsp_1; 1.
DR   PROSITE; PS50092; TSP1; 1.
DR   SMART; SM00209; TSP1; 1.
KW   Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    330       MINDIN.
SQ   SEQUENCE   330 AA;  36014 MW;  ECBCF07A0345A83A CRC64;


  Query Match             86.6%;  Score 1520.5;  DB 11;  Length 330;
  Best Local Similarity   86.4%;  Pred. No. 1.3e-120;
  Matches  287;  Conservative   17;  Mismatches   25;  Indels    3;  Gaps    2;

Qy       1 MENPSPAAALGKALCALLLATLGA-AGQPLGGESICSARALAKYSITFTGKWSQTAFPKQ 59
           ||| |   :| : |   ||| ||: |||||||||:|:|| ||:||||||||||||||||
Db       1 MENVS--FSLDRTLWVFLLAMLGSTAGQPLGGESVCTARPLARYSITFTGKWSQTAFPKQ 58

Qy      60 YPLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQS 119
           ||||||||||||||||||||||||||:|||||||||||||||||||||||||| |||
Db      59 YPLFRPPAQWSSLLGAAHSSDYSMWRKNEYVSNGLRDFAERGEAWALMKEIEAAGEKLQS 118

Qy     120 VHAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQ 179
           |||||||||||||||||||||  ||||||||||||||||:|||||:| ||:||
Db     119 VHAVFSAPAVPSGTGQTSAELEVHPRHSLVSFVVRIVPSPDWFVGIDSLDLCEGGRWKEQ 178

Qy     180 AALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARV 239
             |||||:|||||||||||||||||||||||||:|||||||||||||:|||||:|
Db     179 VVLDLYPHDAGTDSGFTFSSPNFATIPQDTVTEITASSPSHPANSFYYPRLKSLPPIAKV 238

Qy     240 TLVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRLGTK 299
           | |||||||||| ||:  | || |||||| |||||||||||||||||||| ||:|| |
Db     239 TFVRLRQSPRAFAPPSLDLASRGNEIVDSLSVPETPLDCEVSLWSSWGLCGGPCGKLGAK 298

Qy     300 SRTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
           |||||||||||||:|||||||||| |||||
Db     299 SRTRYVRVQPANNGTPCPELEEEAECAPDNCV 330


RESULT   4
O42112
ID   O42112      PRELIMINARY;      PRT;   331 AA.
AC   O42112;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MINDIN2.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Rasborinae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Higashijima S., Nose A., Eguchi G., Hotta Y., Okamoto H.;
RL   Dev. Biol. 0:0-0(1997).
DR   EMBL; AB006085; BAA22809.1; -.
```

```
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00090; tsp_1; 1.
DR   PROSITE; PS50092; TSP1; 1.
DR   SMART; SM00209; TSP1; 1.
SQ   SEQUENCE   331 AA;  36887 MW;  1D95D82B6549D273 CRC64;

  Query Match           63.1%;  Score 1108;  DB 13;  Length 331;
  Best Local Similarity   61.6%;  Pred. No. 9.3e-86;
  Matches  197;  Conservative   51;  Mismatches   64;  Indels    8;  Gaps    3;

Qy      17 LLLATL-GAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQWSSLLGA 75
           : || | |    |:  : :|:| : ||| :||||:|:|||||| |||:||||||| |:|
Db      15 MTLALLSGVPAMPVDVDRMCTAPSTAKYRLTFTGQWTQTAFPKHYPLYRPPAQWSPLIGV 74

Qy      76 AHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSVHAVFSAPAVPSGTGQ 135
            ||||| :|::|:| |||:|:|:|| ||| |:||:|||| :|||: :|||||| :|||
Db      75 THSSDYHLWQRNEYASNGVREFSERAEAWTLIKEVEAAGERIQSVYGLFSAPAVVAGTGH 134

Qy     136 TSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQAALDLYPYDAGTDSGF 195
           : | ||  ||||:||:||||||||||:|||||:||:||  :|  :|:|||||||||||||
Db     135 ATTEFEVFARHSLLSFIVRIVPSPDWFIGVDSLNLCEGDHGKENISLELYPYDAGTDSGF 194

Qy     196 TFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVTLVRLRQSPRAFIPPA 255
           ||||||| ||||| ||:|||| |||||||||||||| |||||:|:| ::: :    |
Db     195 TFSSPNFETIPQDKVTQITSSFPSHPANSFYYPRLKHLPPIAKVSLTKIKNNQ---IFSL 251

Qy     256 PVLPSRDNEIVDSASVP----ETPLDCEVSLWSSWGLCGGHCGRLGTKSRTRYVRVQPAN 311
           |: |:: |:|     :      ||||||||:|| |||| | ||  | | |||: : |||
Db     252 PIQPTQSNQIPSGNEIDGPLINTPLDCEVSVWSPWGLCKGQCGEKGVKHRTRYIHMHPAN 311

Qy     312 NGSPCPELEEEAECVPDNCV 331
           ||:||| |||:  |:|||||
Db     312 NGAPCPSLEEKRLCIPDNCV 331


RESULT  5
O42111
ID   O42111      PRELIMINARY;      PRT;   334 AA.
AC   O42111;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MINDIN1.
GN   MINDIN2 OR MINDIN1.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Rasborinae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Higashijima S., Nose A., Eguchi G., Hotta Y., Okamoto H.;
RL   Dev. Biol. 0:0-0(1997).
DR   EMBL; AB006084; BAA22808.1; -.
DR   ZFIN; ZDB-GENE-990415-160; mindin1.
DR   ZFIN; ZDB-GENE-990415-161; mindin2.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00090; tsp_1; 1.
DR   PROSITE; PS50092; TSP1; 1.
DR   SMART; SM00209; TSP1; 1.
SQ   SEQUENCE   334 AA;  37233 MW;  7451BF2F95AEDF05 CRC64;

  Query Match           49.4%;  Score 868;  DB 13;  Length 334;
  Best Local Similarity   49.7%;  Pred. No. 1.8e-65;
  Matches  166;  Conservative   51;  Mismatches   91;  Indels   26;  Gaps    6;

Qy      10 LGKALCALLLATLGAA---GQPLGGESICSARALAKYSITFTGKWSQTAFPKQYPLFRPP 66
           | : |  ||  ||  |       | |  ||||  | | : ||| ||   ||||||||||
Db      12 LQQLLVVLLRFTLSCAALVNSTNGTE--CSARGPASYIVVFTGHWSPQTFPKQYPLFRPP 69

Qy      67 AQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSVHAVFSA 126
           |||| |:   |:  | :|::    |:|:: |||:|    |:|: : | :  :|| :::
Db      70 AQWSKLMVVTHNEQYRLWQEGAPASDGMKSFAEQGLTVDLVKDAKEARKR-RSVGSMYRT 128

Qy     127 PAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQAALDLYP 186
           :||| | :| |: :  |  ||| :|:::|||||||||| |:||:| :|:::   ||:|
Db     129 AGIPSGIGHSSTEVLLTPRSPLVSLIVKLIPSPDWFVGVDGLNLCEGGKWKQEVTFDLHP 188

Qy     187 YDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVTLVR--- 243
           :|||||||||||||| | | : :| ||| |:|||||||||| |||:| : : |
Db     189 FDAGTDSGFTFSSPNFPTTPPENITMITSQKPNHPANSFYYPRLNELPPLATIWVKRQSR 248

Qy     244 --LRQSPRAFIPPAPVLPSRDNEIVDSASVP----ETPLDCEVSLWSSWGLCGGHCGRLG 297
             :|| |                | |:  || |    |||||||||:|||||| | | | |
Db     249 LPVRQQNRL-----------SNHILPDASKPHRFSETPLDCEVSMWSSWGLCFGPCARGG 297

Qy     298 TKSRTRYVRVQPANNGSPCPELEEEAECVPDNCV 331
           : ||||: ::|||:|||||||||: || | ||:
Db     298 LRHRTRYILLKPANSGSPCPELEEQEECTPHNCL 331


RESULT  6
Q9V746
ID   Q9V746      PRELIMINARY;      PRT;   601 AA.
AC   Q9V746;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MSPO PROTEIN.
GN   MSPO OR CG10145.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
```

```
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003813; AAF58219.1; -.
DR   FlyBase; FBgn0020269; mspo.
DR   InterPro; IPR002106; AA_tRNA_ligase_II.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00090; tsp_1; 1.
DR   SMART; SM00209; TSP1; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS50092; TSP1; 1.
SQ   SEQUENCE   601 AA;  65434 MW;  529EEDCD348AC8EF CRC64;

  Query Match           30.1%;  Score 528.5;  DB 5;  Length 601;
  Best Local Similarity  27.2%;  Pred. No. 1.8e-36;
  Matches  138;  Conservative   53;  Mismatches  132;  Indels  185;  Gaps   12;

Qy       2 ENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQYP 61
           ::||| : |        |||   | ||   :: |:   || | :     |::  ||| ||
Db      86 DDPSPVSNLNPTAS---LAT-PPATQP-APQTGCTLDRLAVYKVVLHTYWTRELFPKHYP 140

Qy      62 LFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEA--------------WALM 107
             :|| |||:  ||  |:::|:::   |  :  :: ||| |                 |
Db     141 DWRPTAQWTKTLGRTHNANYALYHIGQPATAAVKQFAESGRTDLLDSNAGEQQQVQMQLQ 200

Qy     108 KEIEAAGE-----------------------------------------ALQSVHAVFS 125
           :::|                                             | :||   ||
Db     201 SQMQAGKSPSGGISSGTTSFNATAASTATPTGGSGGSGGSGGSGGTGTTTAERSVFDEFS 260

Qy     126 APAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQAALDLY 185
           ||:| | |:: |:: |   ||||| : |||||||||:|||| :|| |  | :   ::|
Db     261 MPAIPMGAGRSEAKVFVDSNHSLVSLMTRIVPSPDWFIGVDSFELCVGGSWIDTVTVELD 320

Qy     186 PYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVTLVRL- 244
           | |||||:||||::||: | ||  :  |||  | ||| |||||: | |||||    ::|
Db     321 PLDAGTDNGFTFTAPNWPTEPQGVIYRITSRYPGHPAGSFYYPKSKRLPPIATFQFIKLK 380

Qy     245 ------------------------------------------------------------ 244

Db     381 EYELSEVFNIAEDDRKYETVQTQTHLDAEHNHVEMNNELSASIERERQTEQQQLQQNDDE 440

Qy     245 RQSPRA--------------FIPPAP-----VLPSRDNEIV---------------DSAS 270
           ||  |:               : |||     |:|  |   :                |: :
Db     441 RQRIRSQLLAKMNPIYGSNNSLQPAPGQVVSVVPKNDKHAILQSIASSYRRAADASDANA 500

Qy     271 VPETP--------------------------LDCEVSLWSSWGLCGGHCGRLGTKSRTR 303
             ||                           || || || |  |   || :|   | |
Db     501 SKPTPSAIGGGKAGGAVGGGAATRRRSSAQRRRDCRVSHWSEWTACSKSCG-VGEMHRYR 559

Qy     304 YVRVQPANNGSPCPELEEEAECVPD-NC 330
           |     |  || |::   |  : ||
Db     560 KVIKHGKRGGRQCPALQQSKWCGTERNC 587


RESULT   7
O02029
ID   O02029      PRELIMINARY;      PRT;   598 AA.
AC   O02029;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   M-SPONDIN.
GN   MSPO OR CG10145.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON S;
RA   Umemiya T., Takeichi M., Nose A.;
RL   Dev. Biol. 0:0-0(0).
DR   EMBL; AF000178; AAB58341.1; -.
DR   FlyBase; FBgn0020269; mspo.
DR   InterPro; IPR002106; AA_tRNA_ligase_II.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00090; tsp_1; 1.
DR   SMART; SM00209; TSP1; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS50092; TSP1; 1.
SQ   SEQUENCE   598 AA;  65251 MW;  02973A52C514350D CRC64;

  Query Match           29.5%;  Score 518;  DB 5;  Length 598;
  Best Local Similarity  26.5%;  Pred. No. 1.4e-35;
  Matches  134;  Conservative   51;  Mismatches  138;  Indels  182;  Gaps   11;

Qy       2 ENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQYP 61
           ::|||   :    | |         |  |   |:   || | :     |::  ||| ||
Db      86 DDPSPVPNMNPT--ASLATPPATQTAPQTG---CTLDRLAVYKVVLHTYWTRELFPKHYP 140

Qy      62 LFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEA--------------WALM 107
             :|| |||:  ||  |:::|:::   |  :  :: ||| |                 |
Db     141 DWRPTAQWTKTLGRTHNANYALYHIGQPATAAVKQFAESGRTDLLDSNAGEQQQVQMQLQ 200

Qy     108 KEIEAAGE-----------------------------------------ALQSVHAVFSAPA 128
           :::|                                             | :||   || ||
Db     201 SQMQAGKSPSGGISSGTTSFNATAASTATPTGGSGGSGGSGVTGTTTAERSVFDEFSMPA 260

Qy     129 VPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQAALDLYPYD 188
           :| | |:: |:: |   ||||| : ||||||||||:|||| :|| |  | :   ::| | |
Db     261 IPMGAGRSEAQVFVDSNHSLVSLMTRIVPSPDWFIGVDSFELCVGGSWIDTVTVELDPLD 320

Qy     189 AGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVTLVRL---- 244
           ||||:|||::||: | ||  :  |||  | ||| |||||: | |||||    ::|
Db     321 AGTDNGFTFTAPNWPTAPQGVIYRITSRYPGHPAGSFYYPKSKRLPPIATFQFIKLKEYE 380

Qy     245 ------------------------------------------------------------RQS 247
                                                                        ||
Db     381 LSEVFNIAEDDRKYETVQTQTHLDAEHNHVEMNNELSASIERERQTEQQQLQQNDDERQR 440

Qy     248 PRA--------------FIPPAP-----VLPSRDNEIV---------------DSASVPE 273
           |:               :  ||      |:|  |   :                |: :
Db     441 IRSQLLAKMNPIYGSNNSLQAAPGQVVSVVPKNDKHAILQSIASSYRRAADASDANASKP 500

Qy     274 TP--------------------------LDCEVSLWSSWGLCGGHCGRLGTKSRTRYVR 306
           ||                           || || || |  |   || :|   | | |
Db     501 TPSAIGGGKAGGAVGGGAATRRRSSAQRRRDCRVSHWSEWTACSKSCG-VGEMHRYRKVI 559

Qy     307 VQPANNGSPCPELEEEAECVPD-NC 330
             |  || |::   |  : ||
Db     560 KHGKRGGRQCPALQQSKWCGTERNC 584


RESULT   8
O42113
ID   O42113      PRELIMINARY;      PRT;   808 AA.
AC   O42113;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   F-SPONDIN1.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Rasborinae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Higashijima S., Nose A., Eguchi G., Hotta Y., Okamoto H.;
RL   Dev. Biol. 0:0-0(1997).
DR   EMBL; AB006086; BAA22810.1; -.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR002861; Reeler.
```

```
DR   Pfam; PF00090; tsp_1; 6.
DR   Pfam; PF02014; Reeler; 1.
DR   PROSITE; PS50092; TSP1; 6.
DR   SMART; SM00209; TSP1; 6.
SQ   SEQUENCE   808 AA;  90645 MW;  0A24154AA4A89EC7 CRC64;


  Query Match             26.8%;  Score 470.5;  DB 13;  Length 808;
  Best Local Similarity   33.2%;  Pred. No. 2.1e-31;
  Matches  114;  Conservative   52;  Mismatches  132;  Indels   45;  Gaps   11;

Qy       9 ALGKALCALLLATLGAAGQPLGGESI------CSARALAKYSITFTGKWSQTAFPKQYPL 62
           :| | :|           : | ||:       | |   ||| :|| | ||:   || ||
Db     176 SLTKRMC---------EKESLYGETTDKPLLDCCACGTAKYRVTFYGNWSEKLHPKDYP- 225

Qy      63 FRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSVHA 122
            |    ||:|:||:|| :| :|    | | |::  || |    : :||   |: : :|
Db     226 -RRANHWSALIGASHSKNYILWEYGGYASEGVKQVAELGSPVKMEEEIRQKGDEVLTVIK 284

Qy     123 V---------FSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDG 173
           :         :  | |      |||  | |   |:||:  : ||||| ||: | |||
Db     285 MKAQWPAWQPLNVRAAP------SAEFSVDRTRHLMSFLTMLGPSPDWNVGLSSEDLCTR 338

Qy     174 D-RWREQAALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKA 232
           :  | ::   || |:|||||| :: ||| :| ||: :  :||    || : || |   |
Db     339 ECGWVQKVVQDLIPWDAGTDSGVSYESPNKSTAPQEKIRPLTSL--DHPQSPFYDPEGGA 396

Qy     233 LPPIARVTLVRL-RQSPRAFIPPAPVLPSRDNEIVDSA----SVPETPLDCEVSLWSSWG 287
           : |:||: : |: |:  :  | |  |    |: : | |        :||  |  | ||
Db     397 ITPVARLVVERIARKGEQCNIVPDNV----DDIVADIAQEEKEEDDTPETCIYSNWSPWS 452

Qy     288 LCGGHCGRLGTKSRTRYVRVQPANNGSPCPELEEEAECVPDNC 330
           |       | : | | :: |  :   |||  ::   |:   |
Db     453 ACSSSTCEKGKRMRQRMLKAQ-LDLSVPCPHTQDFEPCIGPGC 494


RESULT  9
Q9W770
ID   Q9W770      PRELIMINARY;      PRT;   802 AA.
AC   Q9W770;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   F-SPONDIN PRECURSOR.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99211385; PubMed=10197528;
RA   Debby-Brafman A., Burstyn-Cohen T., Klar A., Kalcheim C.;
RT   "F-Spondin, expressed in somite regions avoided by neural crest cells,
RT   mediates inhibition of distinct somite domains to neural crest
RT   migration.";
RL   Neuron 22:475-488(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99325518; PubMed=10399931;
RA   Burstyn-Cohen T., Tzarfaty V., Frumkin A., Feinstein Y., Stoeckli E.,
RA   Klar A.;
RT   "F-spondin is required for accurate pathfinding of commissural axons
RT   at the floor plate.";
RL   Neuron 23:233-246(1999).
DR   EMBL; AF149302; AAD41495.1; -.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR002861; Reeler.
DR   Pfam; PF00090; tsp_1; 6.
DR   Pfam; PF02014; Reeler; 1.
DR   PROSITE; PS50092; TSP1; 5.
DR   SMART; SM00209; TSP1; 6.
KW   Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    802       F-SPONDIN.
SQ   SEQUENCE   802 AA;  90509 MW;  0644D2BDD0A0FE12 CRC64;


  Query Match             26.5%;  Score 464.5;  DB 13;  Length 802;
  Best Local Similarity   33.2%;  Pred. No. 6.8e-31;
  Matches  111;  Conservative   52;  Mismatches  144;  Indels   27;  Gaps   10;

Qy       9 ALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQ 68
           :| | :|    |: |   :|      | |   ||| :|| | ||:   || :|  |
Db     171 SLTKRICEQDSASEGVTDKP---TLDCCACGTAKYRLTFYGNWSEKTHPKDFP--RRTNH 225

Qy      69 WSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEI-EAAGEALQSVHAVFSAP 127
           ||:::|::|| :| :|    | | |::  || |    : :|| : : | |  : |    |
Db     226 WSAIIGSSHSKNYILWEYGGYASEGVKQVAELGSPVKMEEEIRQQSDEVLTVIKAKAQWP 285

Qy     128 AVP--SGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGD-RWREQAALDL 184
           |    :     |||  | |   |:||:  : |||| ||: : |||  |  | ::   ||
Db     286 AWQPLNVRAAPSAEFSVDRHRHLMSFLTMLGPSPDWNVGLSAEDLCTKDCGWVQKVVQDL 345

Qy     185 YPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVTLVRL 244
           |:|||||| |: |||  |:||: :  :||    || : || |   ::  :||| | |:
Db     346 IPWDAGTDSGVTYESPNKPTVPQEKIRPLTSL--DHPQSPFYDPEGGSIKLVARVVLERI 403

Qy     245 RQSPRA--FIPP------APVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCGRL 296
           :      |:|       | : |    |        :||  |  | || |  |
Db     404 ARKGEQCNFVPDNIDDIVADLAPEEKEE-------DDTPETCIYSNWSPWSACSSSTCEK 456

Qy     297 GTKSRTRYVRVQPANNGSPCPELEEEAECVPDNC 330
           | : | | :: |  :   |||: ::   |:   |
Db     457 GKRMRQRMLKAQ-LDLSVPCPDTQDFQPCMGPGC 489


RESULT  10
Q9HCB6
ID   Q9HCB6      PRELIMINARY;      PRT;   807 AA.
AC   Q9HCB6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VSGP/F-SPONDIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Miyamoto K., Morishita Y., Yamazaki M., Minamino N., Kangawa K.,
RA   Matsuo H., Mizutani T., Yamada K., Minegishi T.;
RT   "Isolation and characterization of vascular smooth muscle cell growth
RT   promoting factor from bovine ovarian follicular fluid, and its cDNA
RT   cloning from bovine and human ovary.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB051390; BAB18461.1; -.
DR   InterPro; IPR002861; Reeler.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF02014; Reeler; 1.
DR   Pfam; PF00090; tsp_1; 6.
DR   SMART; SM00209; TSP1; 6.
DR   PROSITE; PS50092; TSP1; 5.
SQ   SEQUENCE   807 AA;  90987 MW;  06FC52484206BB55 CRC64;


  Query Match             26.2%;  Score 460.5;  DB 4;  Length 807;
  Best Local Similarity   34.5%;  Pred. No. 1.5e-30;
  Matches  115;  Conservative   50;  Mismatches  143;  Indels   25;  Gaps   11;

Qy       9 ALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQ 68
           :| | ||       |   :|:      | |   ||| :|| | ||:   || ||  |
Db     176 SLTKKLCEQDSTFDGVTDKPI---LDCCACGTAKYRLTFYGNWSEKTHPKDYP--RRANH 230

Qy      59 WSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEI-EAAGEALQSVHAVFSAP 127
```

```
          ||:::| :|| :| :|    | | |::  || |    : :|| : : | |  : |    |
Db    231 WSAIIGGSHSKNYVLWEYGGYASEGVKQVAELGSPVKMEEEIRQQSDEVLTVIKAKAQWP 290

Qy    128 AVP--SGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGD-RWREQAALDL 184
          |    :     |||  | |   |:||:  : ||||| ||: : |||  :  | ::   ||
Db    291 AWQPLNVRAAPSAEFSVDRTRHLMSFLTMMGPSPDWNVGLSAEDLCTKECGWVQKVVQDL 350

Qy    185 YPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVTLVRL 244
          |:|||||||| |: |||  ||||: :  :||    || : || |   ::  :||| : |:
Db    351 IPWDAGTDSGVTYESPNKPTIPQEKIRPLTSL--DHPQSPFYDPEGGSITQVARVVIERI 408

Qy    245 -RQSPRAFIPPAPVLPSRDNEIVDSASVPE------TPLDCEVSLWSSWGLCGGHCGRLG 297
           |:  :  | |  |    |: : | |  ||      ||  |  | || |  |       |
Db    409 ARKGEQCNIVPDNV----DDIVADLA--PEEKDEDDTPETCIYSNWSPWSACSSSTCDKG 462

Qy    298 TKSRTRYVRVQPANNGSPCPELEEEAECVPDNC 330
          : | | :: |  :   |||: ::   |:   |
Db    463 KRMRQRMLKAQ-LDLSVPCPDTQDFQPCMGPGC 494
```

RESULT 11
Q9GLX9
ID Q9GLX9 PRELIMINARY; PRT; 807 AA.
AC Q9GLX9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE VSGP/F-SPONDIN.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Miyamoto K., Morishita Y., Yamazaki M., Minamino N., Kangawa K.,
RA Matsuo H., Mizutani T., Yamada K., Minegishi T.;
RT "Isolation and characterization of vascular smooth muscle cell growth
RT promoting factor from bovine ovarian follicular fluid, and its cDNA
RT cloning from bovine and human ovary.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB051389; BAB18460.1; -.
DR InterPro; IPR002861; Reeler.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF02014; Reeler; 1.
DR Pfam; PF00090; tsp_1; 6.
DR SMART; SM00209; TSP1; 6.
DR PROSITE; PS50092; TSP1; 5.
SQ SEQUENCE 807 AA; 90976 MW; 4C484B331FB1034C CRC64;

Query Match 26.2%; Score 460.5; DB 6; Length 807;
Best Local Similarity 34.5%; Pred. No. 1.5e-30;
Matches 115; Conservative 50; Mismatches 143; Indels 25; Gaps 11;

```
Qy      9 ALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQ 68
          :| | ||       |   :|:     | |   ||| :|| | ||:   || ||  |
Db    176 SLTKKLCEQDSTFDGVTDKPI---LDCCACGTAKYRLTFYGNWSEKTHPKDYP--RRANH 230

Qy     69 WSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEI-EAAGEALQSVHAVFSAP 127
          ||:::| :|| :| :|    | | |::  || |    : :|| : : | |  : |
Db    231 WSAIIGGSHSKNYVLWEYGGYASEGVKQVAELGSPVKMEEEIRQQSDEVLTVIKAKAQWP 290

Qy    128 AVP--SGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGD-RWREQAALDL 184
          |    :     |||  | |   |:||:  : ||||| ||: : |||  :  | ::   ||
Db    291 AWQPLNVRAAPSAEFSVDRTRHLMSFLTMMGPSPDWNVGLSAEDLCTKECGWVQKVVQDL 350

Qy    185 YPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIARVTLVRL 244
          |:|||||||| |: |||  ||||: :  :||    || : || |   ::  :||| : |:
Db    351 IPWDAGTDSGVTYESPNKPTIPQEKIRPLTSL--DHPQSPFYDPEGGSITQVARVVIERI 408

Qy    245 -RQSPRAFIPPAPVLPSRDNEIVDSASVPE------TPLDCEVSLWSSWGLCGGHCGRLG 297
           |:  :  | |  |    |: : | |  ||      ||  |  | || |  |       |
Db    409 ARKGEQCNIVPDNV----DDIVADLA--PEEKDEDDTPETCIYSNWSPWSACSSSTCDKG 462

Qy    298 TKSRTRYVRVQPANNGSPCPELEEEAECVPDNC 330
          : | | :: |  :   |||: ::   |:   |
Db    463 KRMRQRMLKAQ-LDLSVPCPDTQDFQPCMGPGC 494
```

RESULT 12
O94862
ID O94862 PRELIMINARY; PRT; 624 AA.
AC O94862;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE KIAA0762 PROTEIN (FRAGMENT).
GN KIAA0762.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:277-286(1998).
DR EMBL; AB018305; BAA34482.1; -.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp_1; 6.
DR PROSITE; PS50092; TSP1; 5.
DR SMART; SM00209; TSP1; 6.
FT NON_TER 1 1
SQ SEQUENCE 624 AA; 70557 MW; 40F2238D29024D03 CRC64;

Query Match 26.0%; Score 457; DB 4; Length 624;
Best Local Similarity 35.5%; Pred. No. 2.1e-30;
Matches 109; Conservative 47; Mismatches 129; Indels 22; Gaps 10;

```
Qy     35 CSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGL 94
          | |   ||| :|| | ||:   || ||  |    ||:::| :|| :| :|    | | |:
Db     16 CCACGTAKYRLTFYGNWSEKTHPKDYP--RRANHWSAIIGGSHSKNYVLWEYGGYASEGV 73

Qy     95 RDFAERGEAWALMKEI-EAAGEALQSVHAVFSAPAVP--SGTGQTSAELEVQRRHSLVSF 151
          :  || |    : :|| : : | |  : |    ||    :     |||  | |   |:||
Db     74 KQVAELGSPVKMEEEIRQQSDEVLTVIKAKAQWPAWQPLNVRAAPSAEFSVDRTRHLMSF 133

Qy    152 VVRIVPSPDWFVGVDSLDLCDGD-RWREQAALDLYPYDAGTDSGFTFSSPNFATIPQDTV 210
          :  : ||||| ||: : |||  :  | ::   || |:|||||||| |: |||  |||: :
Db    134 LTMMGPSPDWNVGLSAEDLCTKECGWVQKVVQDLIPWDAGTDSGVTYESPNKPTIPQEKI 193

Qy    211 TEITSSSPSHPANSFYYPRLKALPPIARVTLVRL-RQSPRAFIPPAPVLPSRDNEIVDSA 269
            :||    || : || |   ::  :||| : |: |:  :  | |  |    |: : | |
Db    194 RPLTSL--DHPQSPFYDPEGGSITQVARVVIERIARKGEQCNIVPDNV----DDIVADLA 247

Qy    270 SVPE------TPLDCEVSLWSSWGLCGGHCGRLGTKSRTRYVRVQPANNGSPCPELEEEA 323
            ||      ||  | | || | |       | : | | :: |  :   |||: ::
Db    248 --PEEKDEDDTPETCIYSNWSPWSACSSSTCDKGKRMRQRMLKAQ-LDLSVPCPDTQDFQ 304

Qy    324 ECVPDNC 330
           |:  |
Db    305 PCMGPGC 311
```

RESULT 13
O76822
ID O76822 PRELIMINARY; PRT; 898 AA.
AC O76822;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)

```
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   F-SPONDIN.
GN   F-SPONDIN.
OS   Branchiostoma floridae (Florida lancelet) (Amphioxus).
OC   Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC   Branchiostoma.
OX   NCBI_TaxID=7739;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shimeld S.M.;
RT   "Characterisation of AmphiF-spondin reveals the modular evolution of
RT   chordate F-spondin genes.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ006096; CAA06854.1; -.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR002861; Reeler.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF02014; Reeler; 1.
DR   Pfam; PF00090; tsp_1; 3.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00209; TSP1; 3.
DR   PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR   PROSITE; PS50092; TSP1; 3.
KW   Repeat.
SQ   SEQUENCE   898 AA;  100418 MW;  E1DE840C7B9C266A CRC64;

  Query Match             26.0%;  Score 457;  DB 5;  Length 898;
  Best Local Similarity   25.9%;  Pred. No. 3.4e-30;
  Matches  113;  Conservative   55;  Mismatches  115;  Indels  154;  Gaps   10;

Qy     35 CSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGL 94
Db    189 CTACGAGKYRMTFYGLWSQQTHPKDYPKY--GTHWSAIIGATHSKDYTIWEYGGYSTEGV 246

Qy     95 RDFAERGEAWALMKEIEAAGEALQSVHAVFSAPAVPSG-----TGQTSAELEVQRRHSLV 149
Db    247 KQVAEWGSPTKLEREMKEQGNKIRTV--VKTDPQWPAYEPRNIKAPVSSDFGVDSEHHLL 304

Qy    150 SFVVRIVPSPDWFVGVDSLDLCDGDRWREQAALDLYPYDAGTDSGFTFSSPNFATIPQDT 209
Db    305 SFLTMLGPSPDWNVGLSNFDVCQGDQWLDRTEMDLYPWDAGTDSGITYESANSPTIPREK 364

Qy    210 VTEITSSSPSHPANSFYYPRLKALPPIARVTLVRL------------------------- 244
Db    365 IRPITSQDPKNPASPWYSARGKSFPPVARIVIERTGIMDNGCHGKTSRDDEEDIDIFSEE 424

Qy    245 ------------RQSPRA------------------------------------------ 250
Db    425 ISLSKAEKKERKRNKDRNRGRDKGRDRHGHDHDHDGSHKLTKGDLTISMGTMAVNSTAVK 484

Qy    251 --FIPPAPVLPS------------------------RDNEIVDSASVPETP--------- 275
Db    485 IWWAPPSNVHSSILGYLVYYQEGQKLLGPNVLEIPDGETTQIVLGGLEPESPYTIRVVAY 544

Qy    276 ------------------------LDCEVSLWSSWGLCGGHCGRLGTKSRTRYVRVQPAN 311
Db    545 TMEQSNIQSEPAMVTTLPDTSIVDVPCLYSPWSEWSDCSVQCGK-GVRKRNRVLKSLP-- 601

Qy    312 NGSP--CPELEEEAECV 326
Db    602 --SPMCTDELEQTEQCM 616


RESULT  14
Q9XZD0
ID   Q9XZD0      PRELIMINARY;      PRT;   763 AA.
AC   Q9XZD0;

DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   FAT-SPONDIN PROTEIN.
GN   FAT-SPONDIN OR CG6953.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RA   DaRocha S., Baumgartner S.;
RT   "Fat-Spondin, a Drosophila member of the Spondin family, is highly
RT   expressed in fat body and hemocytes.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO PANCREATIC TRYPSIN INHIBITOR (KUNITZ) DOMAIN.
DR   EMBL; AE003804; AAF57910.1; -.
DR   EMBL; AF135119; AAD31715.1; -.
DR   HSSP; P00974; 1BTI.
DR   FlyBase; FBgn0026721; fat-spondin.
DR   InterPro; IPR002223; Kunitz_BPTI.
DR   InterPro; IPR002861; Reeler.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00014; Kunitz_BPTI; 1.
DR   Pfam; PF02014; Reeler; 1.
DR   Pfam; PF00090; tsp_1; 4.
DR   PRINTS; PR00759; BASICPTASE.
DR   SMART; SM00131; KU; 1.
DR   SMART; SM00209; TSP1; 4.
```

```
DR   PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR   PROSITE; PS50279; BPTI_KUNITZ_2; 1.
DR   PROSITE; PS50092; TSP1; 4.
KW   Serine protease inhibitor.
SQ   SEQUENCE   763 AA;  84945 MW;  3292DEDD2CFE4DAB CRC64;


  Query Match             25.4%;  Score 445.5;  DB 5;  Length 763;
  Best Local Similarity   33.3%;  Pred. No. 2.6e-29;
  Matches  111;  Conservative   45;  Mismatches  138;  Indels   39;  Gaps    9;

Qy       2 ENPSPAAALGKALCALLLATLGAAGQPLGGESICSARALAKYSITFTGKWSQTAFPKQYP 61
Db     173 ERKPDAAAAQKECCA-----------------CDE---AKYSFVFEGIWSNETHPKDYP 211

Qy      62 LFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGEAWALMKEIEAAGEALQSVH 121
Db     212 FAIWLTHFSDVIGASHESNFSFWGENHIATAGFRSLAEWGSPAALETELRANGPKLRT-- 269

Qy     122 AVFSAPAV--PSGTGQTSAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGD-RWRE 178
Db     270 -LIKAAGLWYPNVNQNTSSKFRVDRKHPKVSLVSMFGPSPDWVVGISGLDLCTEDCSWKE 328

Qy     179 QAALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSPSHPANSFYYPRLKALPPIAR 238
Db     329 SMDFDLFPWDAGTDSGISYMSPNSETQPPERMYRITTMYPEDPRAPFYNPKSREMTPLAK 388

Qy     239 VTLVR----LRQSPRAFIPPAPVLPSRDNEIVDSASVPETPLDCEVSLWSSWGLCGGHCG 294
Db     389 LYLRREKIVSRNCDDEF------LQALQLEVSDDAEEQDTRAECRVGDYSAWSPCSVSCG 442

Qy     295 RLGTKSRTR-YVRVQPANNGSPCPELEEEAECV 326
Db     443 K-GIRMRSRQYLYPAAADQNKCARQLVAKEMCV 474


RESULT  15
O42114
ID   O42114      PRELIMINARY;      PRT;   803 AA.
AC   O42114;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   F-SPONDIN2.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Rasborinae; Danio.
     NCBI_TaxID=7955;
     [1]
RN   SEQUENCE FROM N.A.
RA   Higashijima S., Nose A., Eguchi G., Hotta Y., Okamoto H.;
RL   Dev. Biol. 0:0-0(1997).
DR   EMBL; AB006087; BAA22811.1; -.
DR   InterPro; IPR000762; PTN_MK.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR002861; Reeler.
DR   Pfam; PF00090; tsp_1; 6.
DR   Pfam; PF02014; Reeler; 1.
DR   PROSITE; PS50092; TSP1; 6.
DR   SMART; SM00193; PTN; 1.
DR   SMART; SM00209; TSP1; 6.
SQ   SEQUENCE   803 AA;  90331 MW;  619C8F54B40DC158 CRC64;


  Query Match             25.2%;  Score 443;  DB 13;  Length 803;
  Best Local Similarity   30.6%;  Pred. No. 4.5e-29;
  Matches  113;  Conservative   47;  Mismatches  131;  Indels   78;  Gaps   10;

Qy      35 CSARALAKYSITFTGKWSQTAFPKQYPLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGL 94
Db     194 CCACGTAKYRVTFYGNWSEKLDPKDYP--RRANHWSALIGASHTRSYVLWEYGGFASEGV 251

Qy      95 RDFAERGEAWALMKEIEAAGEALQSV---HAVFSAPAVPSGTGQTSAELEVQRRHSLVSF 151
Db     252 KQVSEYGSPVKMEEEIRQKGDDVMTVIKTKAQWPAWQPLNIRAAPSAEFSVDRVRHLISF 311

Qy     152 VVRIVPSPDWFVGVDSLDLCDGD-RWREQAALDLYPYDAGTDSGFTFSSPNFATIPQDTV 210
Db     312 LTMLGPSPDWNVGLSGEDLCTRDCGWVQKLVKDLVPWDAGTDSGVSYESPNKPSIPQERI 371

Qy     211 TEITSSSPSHPANSFYYPRLKALPPIARVTLVRLRQ---------------------- 246
Db     372 RPLTSL--DHPQSPFYDPSGGPMTPLGRVVVERIARKGEQCNTVPDTIDDIVADIAQEEK 429

Qy     247 -----SPRAFI------------------------------PPAPVLPSRDNE-IVDSA 269
Db     430 EEGDNTPETCIYSNWSPWSACSSSTCDKGRRMRQRMLKAQLDPNLPCLHTQDFEPCMGPG 489

Qy     270 SVPETPLDCEVSLWSSWGLCGGHCGRLGTKSRTRYVRVQPANNGSPCPELEEEAE----- 324
Db     490 CSEEEASTCMMSEWISWSPCSASCG-MGMRSRERYVKQFP-EDGSSCTLPTEETEKCVVN 547

Qy     325 --CVPDNCV 331
Db     548 DDCSPSSCV 556


Search completed: December 10, 2001, 20:02:39
Job time: 2891 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2001, 13:37:26 ; Search time 1640.62 Seconds
(without alignments)
11111.277 Million cell updates/sec

Title: US-09-170-042A-1
Perfect score: 1105
Sequence: 1 cgctgctcctgccgggtgat.....tgcaggcggccgagggcaca 1105

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1105	100.0	1105	6	AR035961	AR035961 Sequence
2	1098.6	99.4	1634	9	BC002707	BC002707 Homo sapi
3	1084.8	98.2	1807	9	AB027466	AB027466 Homo sapi
4	665.6	60.2	2325	10	AF155196	AF155196 Rattus no
5	362.6	32.8	2116	5	AB006085	AB006085 Danio rer
6	343.8	31.1	400	6	AR037874	AR037874 Sequence
7	339	30.7	400	6	AR125230	AR125230 Sequence
c 8	336.2	30.4	197176	2	AC022763	AC022763 Homo sapi
9	300	27.1	506	6	AR035966	AR035966 Sequence
10	279	25.2	4292	9	AK024499	AK024499 Homo sapi
11	277.4	25.1	2820	9	AK026054	AK026054 Homo sapi
12	258	23.3	3381	5	AB006084	AB006084 Danio rer
13	202.6	18.3	316	6	AR035967	AR035967 Sequence
14	202.6	18.3	316	6	AR035968	AR035968 Sequence
15	118.2	10.7	2565	3	DMAF000178	AF000178 Drosophil
16	110.6	10.0	3489	3	BFL6096	AJ006096 Branchios
17	105	9.5	2466	3	AF135119	AF135119 Drosophil
18	81.2	7.3	3146	5	AB006086	AB006086 Danio rer
19	78	7.1	4061	4	AB051389	AB051389 Bos tauru
20	78	7.1	4061	6	E32471	E32471 Novel vascu
c 21	75	6.8	51289	3	AC005717	AC005717 Drosophil
c 22	75	6.8	73721	2	AC020007	AC020007 Drosophil
23	75	6.8	259817	3	AE003804	AE003804 Drosophil
24	71.6	6.5	3079	6	AX093366	AX093366 Sequence
25	71.6	6.5	3079	6	E32470	E32470 Novel vascu
26	71.6	6.5	3079	9	AB051390	AB051390 Homo sapi
27	71.6	6.5	3999	6	AX093359	AX093359 Sequence
28	71.6	6.5	3999	9	AB018305	AB018305 Homo sapi
c 29	70.6	6.4	76217	2	AC020385	AC020385 Drosophil
c 30	70.6	6.4	160903	3	AC007580	AC007580 Drosophil
c 31	70.6	6.4	258778	3	AE003813	AE003813 Drosophil
32	70	6.3	4029	6	AR007441	AR007441 Sequence
33	70	6.3	4029	6	E32472	E32472 Novel vascu
34	70	6.3	4029	10	RATFSAA	M88469 Rattus norv
c 35	67	6.1	20293	1	SCD20	AL392148 Streptomy
c 36	65.4	5.9	10805	2	AC017707	AC017707 Drosophil
37	65.4	5.9	184621	3	AC007440	AC007440 Drosophil
38	65.4	5.9	266133	3	AE003822	AE003822 Drosophil
39	61.6	5.6	3226	5	AF149302	AF149302 Gallus ga
40	61.6	5.6	3226	6	AR007442	AR007442 Sequence
41	61.2	5.5	59843	8	AP000816	AP000816 Oryza sat
42	61.2	5.5	161266	8	AP001278	AP001278 Oryza sat
c 43	60.8	5.5	12095	1	AE005853	AE005853 Caulobact
44	60.8	5.5	167761	2	AP003221	AP003221 Oryza sat
45	60	5.4	3576	5	AB006087	AB006087 Danio rer

ALIGNMENTS

RESULT 1
AR035961
LOCUS AR035961 1105 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5871969.
ACCESSION AR035961
VERSION AR035961.1 GI:5952629
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1105)
AUTHORS Hastings,G. and Dillon,P.J.
TITLE Nucleic acids encoding human neuronal attachment factor-1
JOURNAL Patent: US 5871969-A 1 16-FEB-1999;
FEATURES Location/Qualifiers
source 1. .1105
/organism="unknown"
BASE COUNT 178 a 393 c 362 g 172 t
ORIGIN

```
Query Match             100.0%;  Score 1105;  DB 6;  Length 1105;
Best Local Similarity   100.0%;  Pred. No. 2.2e-150;
Matches 1105;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CGCTGCTCCTGCCGGGTGATGGAAAACCCCAGCCCGGCCGCCGCCCTGGGCAAGGCCCTC 60

Qy      61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGCGCTCTCCTCCTGGCCACTCTCGGCGCCGCCGGCCAGCCTCTTGGGGGAGAGTCCATC 120

Qy     121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TGTTCCGCCAGAGCCCTGGCCAAATACAGCATCACCTTCACGGGCAAGTGGAGCCAGACG 180

Qy     181 gccttccccaagcagtacccccctgttccgccccccctgcccagtggtcttcgctgctgggg 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GCCTTCCCCAAGCAGTACCCCCTGTTCCGCCCCCCTGCCCAGTGGTCTTCGCTGCTGGGG 240

Qy     241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCCGCGCATAGCTCCGACTACAGCATGTGGAGGAAGAACCAGTACGTCAGTAACGGGCTG 300

Qy     301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CGCGACTTTGCGGAGCGCGGCGAGGCCTGGGCGCTGATGAAGGAGATCGAGGCGGCGGGG 360

Qy     361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GAGGCGCTGCAGAGCGTGCACGCGGTGTTTTCGGCGCCCGCCGTCCCCAGCGGCACCGGG 420

Qy     421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CAGACGTCGGCGGAGCTGGAGGTGCAGCGCAGGCACTCGCTGGTCTCGTTTGTGGTGCGC 480

Qy     481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATCGTGCCCAGCCCCGACTGGTTCGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGGAC 540

Qy     541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CGTTGGCGGGAACAGGCGGCGCTGGACCTGTACCCCTACGACGCCGGGACGGACAGCGGC 600

Qy     601 ttcaccttctcctccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TTCACCTTCTCCTCCCCAACTTCGCCACCATCCCGCAGGACACGGTGACCGAGATAACG 660

Qy     661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TCCTCCTCTCCCAGCCACCCGGCCAACTCCTTCTACTACCCGCGGCTGAAGGCCCTGCCT 720

Qy     721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CCCATCGCCAGGGTGACACTGGTGCGGCTGCGACAGAGCCCCAGGGCCTTCATCCCTCCC 780

Qy     781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GCCCCAGTCCTGCCCAGCAGGGACAATGAGATTGTAGACAGCGCCTCAGTTCCAGAAACG 840

Qy     841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CCGCTGGACTGCGAGGTCTCCCTGTGGTCGTCCTGGGGACTGTGCGGAGGCCACTGTGGG 900

Qy     901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 AGGCTCGGGACCAAGAGCAGGACTCGCTACGTCCGGGTCCAGCCCGCCAACAACGGGAGC 960

Qy     961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CCCTGCCCCGAGCTCGAAGAAGAGGCTGAGTGCGTCCCTGATAACTGCGTCTAAGACCAG 1020

Qy    1021 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 AGCCCCGCAGCCCCTGGGGCCCCCCGGAGCCATGGGGTGTCGGGGGCTCCTGTGCAGGCT 1080

Qy    1081 catgctgcaggcggccgagggcaca 1105
           |||||||||||||||||||||||||
Db    1081 CATGCTGCAGGCGGCCGAGGGCACA 1105
```

RESULT 2
BC002707
LOCUS BC002707 1634 bp mRNA PRI 12-JUL-2001
DEFINITION Homo sapiens, spondin 2, extracellular matrix protein, clone MGC:3472 IMAGE:3630467, mRNA, complete cds.
ACCESSION BC002707
VERSION BC002707.1 GI:12803740
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1634)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R., Lim,M., Maduro,Q.L., Masiello,C., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Snyder,B., Stantripop,S., Thomas,P.J., Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 12 Row: a Column: 16
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6912681.

FEATURES Location/Qualifiers
source 1. .1634
/organism="Homo sapiens"
/db_xref="LocusID:10417"
/db_xref="taxon:9606"
/clone="MGC:3472 IMAGE:3630467"
/tissue_type="Placenta, choriocarcinoma"
/clone_lib="NIH_MGC_21"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
CDS 69. .1064
/codon_start=1
/product="spondin 2, extracellular matrix protein"
/protein_id="AAH02707.1"
/db_xref="GI:12803741"
/translation="MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARAPAKYS
ITFTGKWSQTAFPKQYPLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGE
AWALMKEIEAAGEALQSVHEVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPD
WFVGVDSLDLCDGDRWREQAALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSP
SHPANSFYYPRLKALPPIARVTLLRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPL
DCEVSLWSSWGLCGGHCGRLGTKSRTRYVRVQPANNGSPCPELEEEAECVPDNCV"

```
BASE COUNT      293 a    550 c    501 g    290 t
ORIGIN

  Query Match            99.4%;  Score 1098.6;  DB 9;  Length 1634;
  Best Local Similarity  99.6%;  Pred. No. 1.5e-149;
  Matches 1101;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51 CGCTGCTCCTGCCGGGTGATGGAAAACCCCAGCCCGGCCGCCGCCCTGGGCAAGGCCCTC 110

Qy      61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     111 TGCGCTCTCCTCCTGGCCACTCTCGGCGCCGCCGGCCAGCCTCTTGGGGGAGAGTCCATC 170

Qy     121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     171 TGTTCCGCCAGAGCCCCGGCCAAATACAGCATCACCTTCACGGGCAAGTGGAGCCAGACG 230

Qy     181 gccttccccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggg 240
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     231 GCCTTCCCCAAGCAGTACCCCCTGTTCCGCCCCCCTGCGCAGTGGTCTTCGCTGCTGGGG 290

Qy     241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     291 GCCGCGCATAGCTCCGACTACAGCATGTGGAGGAAGAACCAGTACGTCAGTAACGGGCTG 350

Qy     301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     351 CGCGACTTTGCGGAGCGCGGCGAGGCCTGGGCGCTGATGAAGGAGATCGAGGCGGCGGGG 410

Qy     361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     411 GAGGCGCTGCAGAGCGTGCACGAGGTGTTTTCGGCGCCCGCCGTCCCCAGCGGCACCGGG 470

Qy     421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     471 CAGACGTCGGCGGAGCTGGAGGTGCAGCGCAGGCACTCGCTGGTCTCGTTTGTGGTGCGC 530

Qy     481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     531 ATCGTGCCCAGCCCCGACTGGTTCGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGGAC 590

Qy     541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     591 CGTTGGCGGGAACAGGCGGCGCTGGACCTGTACCCCTACGACGCCGGGACGGACAGCGGC 650

Qy     601 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     651 TTCACCTTCTCCTCCCCCAACTTCGCCACCATCCCGCAGGACACGGTGACCGAGATAACG 710

Qy     661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     711 TCCTCCTCTCCCAGCCACCCGGCCAACTCCTTCTACTACCCGCGGCTGAAGGCCCTGCCT 770

Qy     721 cccatcgccagggtgacactggtgcggctgcgacagaggccccagggccttcatccctccc 780
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     771 CCCATCGCCAGGGTGACACTGCTGCGGCTGCGACAGAGCCCCAGGGCCTTCATCCCTCCC 830

Qy     781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     831 GCCCCAGTCCTGCCCAGCAGGGACAATGAGATTGTAGACAGCGCCTCAGTTCCAGAAACG 890

Qy     841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     891 CCGCTGGACTGCGAGGTCTCCCTGTGGTCGTCCTGGGGACTGTGCGGAGGCCACTGTGGG 950

Qy     901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     951 AGGCTCGGGACCAAGAGCAGGACTCGCTACGTCCGGGTCCAGCCCGCCAACAACGGGAGC 1010

Qy     961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1011 CCCTGCCCCGAGCTCGAAGAAGAGGCTGAGTGCGTCCCTGATAACTGCGTCTAAGACCAG 1070

Qy    1021 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1071 AGCCCCGCAGCCCCTGGGGCCCCCCGGAGCCATGGGGTGTCGGGGGCTCCTGTGCAGGCT 1130

Qy    1081 catgctgcaggcggccgagggcaca 1105
           |||||||||||||||||||||||||
Db    1131 CATGCTGCAGGCGGCCGAGGGCACA 1155


RESULT   3
AB027466
LOCUS       AB027466     1807 bp    mRNA            PRI       05-NOV-1999
DEFINITION  Homo sapiens SPON2 mRNA for spondin 2, complete cds.
ACCESSION   AB027466
VERSION     AB027466.1  GI:6172220
KEYWORDS    SPON2; spondin 2.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Manda,R., Kohno,T., Matsuno,Y., Takenoshita,S., Kuwano,H. and
            Yokota,J.
  TITLE     Identification of genes (SPON2 and c20orf2) differentially
            expressed between cancerous and noncancerous lung cells by mRNA
            differential display
  JOURNAL   Genomics 61 (1), 5-14 (1999)
  MEDLINE   99443867
REFERENCE   2  (bases 1 to 1807)
  AUTHORS   Yokota,J., Kohno,T. and Manda,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAY-1999) to the DDBJ/EMBL/GenBank databases. Jun
            Yokota, National Cancer Center Research Institute, Biology
            Division; Tsukiji 5-chome 1-1, Chuo-ku, Tokyo 104-0045, Japan
            (E-mail:jyokota@gan2.ncc.go.jp, Tel:81-3-3547-5272,
            Fax:81-3-3542-0807)
FEATURES             Location/Qualifiers
     source          1. .1807
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4p16.3"
     gene            276. .1271
                     /gene="SPON2"
     CDS             276. .1271
                     /gene="SPON2"
                     /note="alias: DIL-1; extracellular matrix protein"
                     /codon_start=1
                     /product="spondin 2"
                     /protein_id="BAA85892.1"
                     /db_xref="GI:6172221"
                     /translation="MENPSPAAALGKALCALLLATLGAAGQPLGGESICSARAPAKYS
                     ITFTGKWSQTAFPKQYPLFRPPAQWSSLLGAAHSSDYSMWRKNQYVSNGLRDFAERGE
                     AWALMKEIEAAGEALQSVHAVFSAPAVPSGTGQTSAELEVQRRHSLVSFVVRIVPSPD
                     WFVGVDSLDLCDGDRWREQAALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEITSSSP
                     SHPANSFYYPRLKALPPIARVTLVRLRQSPRAFIPPAPVLPSRDNEIVDSASVPETPL
                     DCEVSLWSSWGLCGGHCGRLGTKSRTRYVRVQPANNGSPCPELEEEAECVPDNCV"
BASE COUNT      292 a    621 c    572 g    322 t
ORIGIN

  Query Match            98.2%;  Score 1084.8;  DB 9;  Length 1807;
  Best Local Similarity  99.7%;  Pred. No. 1.4e-147;
  Matches 1097;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     258 CGCTGCTCCTGCCGGGTGATGGAAAACCCCAGCCCGGCCGCCGCCCTGGGCAAGGCCCTC 317

Qy      61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
```

```
Db    318 TGCGCTCTCCTCCTGGCCACTCTCGGCGCCGCCGGCCAGCCTCTTGGGGGAGAGTCCATC 377

Qy    121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
Db    378 TGTTCCGCCAGAGCCCCGGCCAAATACAGCATCACCTTCACGGGCAAGTGGAGCCAGACG 437

Qy    181 gccttccccaagcagtacccccctgttccgccccccctgcccagtggtcttcgctgctgggg 240
Db    438 GCCTTCCCCAAGCAGTACCCCCTGTTCCGCCCCCCTGCGCAGTGGTCTTCGCTGCTGGGG 497

Qy    241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
Db    498 GCCGCGCATAGCTCCGACTACAGCATGTGGAGGAAGAACCAGTACGTCAGTAACGGGCTG 557

Qy    301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
Db    558 CGCGACTTTGCGGAGCGCGGCGAGGCCTGGGCGCTGATGAAGGAGATCGAGGCGGCGGGG 617

Qy    361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
Db    618 GAGGCGCTGCAGAGCGTGCACGCGGTGTTTTCGGCGCCCGCCGTCCCCAGCGGCACCGGG 677

Qy    421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
Db    678 CAGACGTCGGCGGAGCTGGAGGTGCAGCGCAGGCACTCGCTGGTCTCGTTTGTGGTGCGC 737

Qy    481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
Db    738 ATCGTGCCCAGCCCCGACTGGTTCGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGGAC 797

Qy    541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
Db    798 CGTTGGCGGGAACAGGCGGCGCTGGACCTGTACCCCTACGACGCCGGGACGGACAGCGGC 857

Qy    601 ttcaccttctcctccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
Db    858 TTCACCTTCTCCTCCCCCAACTTCGCCACCATCCCGCAGGACACGGTGACCGAGATAACG 917

Qy    661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
Db    918 TCCTCCTCTCCCAGCCACCCGGCCAACTCCTTCTACTACCCGCGGCTGAAGGCCCTGCCT 977

Qy    721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
Db    978 CCCATCGCCAGGGTGACACTGGTGCGGCTGCGACAGAGCCCCAGGGCCTTCATCCCTCCC 1037

Qy    781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
Db   1038 GCCCCAGTCCTGCCCAGCAGGGACAATGAGATTGTAGACAGCGCCTCAGTTCCAGAAACG 1097

Qy    841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
Db   1098 CCGCTGGACTGCGAGGTCTCCCTGTGGTCGTCCTGGGGACTGTGCGGAGGCCACTGTGGG 1157

Qy    901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
Db   1158 AGGCTCGGGACCAAGAGCAGGACTCGCTACGTCCGGGTCCAGCCCGCCAACAACGGGAGC 1217

Qy    961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
Db   1218 CCCTGCCCCGAGCTCGAAGAAGAGGCTGAGTGCGTCCCTGATAACTGCGTCTAAGACCAG 1277

Qy   1021 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
Db   1278 AGCCCCGCAGCCCCTGGGG-CCCCCGGAGCCATGGGGTGTCGGGGGCTCCTGTGCAGGCT 1336

Qy   1081 catgctgcaggcggccgagg 1100
Db   1337 CATGCTGCAGGCGGCCGAGG 1356

RESULT   4
```

```
AF155196
LOCUS       AF155196     2325 bp    mRNA            ROD       23-DEC-1999
DEFINITION  Rattus norvegicus mindin precursor, mRNA, complete cds.
ACCESSION   AF155196
VERSION     AF155196.1  GI:5031505
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2325)
  AUTHORS   Feinstein,Y., Borrell,V., Garcia,C., Burstyn-Cohen,T., Tzarfaty,V.,
            Frumkin,A., Nose,A., Okamoto,H., Higashijima,S., Soriano,A. and
            Klar,A.
  TITLE     F-spondin and mindin: two structurally and functionally related
            genes expressed in the hippocampus that promote outgrowth of
            embryonic hippocampal neurons
  JOURNAL   Development 126 (16), 3637-3648 (1999)
  MEDLINE   99339921
   PUBMED   10409509
REFERENCE   2  (bases 1 to 2325)
  AUTHORS   Feinstein,Y., Borrell,V., Garcia,C., Burstyn-Cohen,T., Tzarfaty,V.,
            Frumkin,A., Nose,A., Okamoto,H., Higashijima,S., Soriano,A. and
            Klar,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-1999) Anatomy & Cell Biology, Hebrew University,
            PO Box 12272, Jerusalem 91120, Israel
FEATURES             Location/Qualifiers
     source          1. .2325
                     /organism="Rattus norvegicus"
                     /strain="Sprague Dawley"
                     /db_xref="taxon:10116"
     CDS             512. .1504
                     /note="adhesion molecule"
                     /codon_start=1
                     /product="mindin precursor"
                     /protein_id="AAD38195.1"
                     /db_xref="GI:5031506"
                     /translation="MENVSFSLDRTLWVFLLAMLGSTAGQPLGGESVCTARPLARYSI
                     TFTGKWSQTAFPKQYPLFRPPAQWSSLLGAAHSSDYSMWRKNEYVSNGLRDFAERGEA
                     WALMKEIEAAGEKLQSVHAVFSAPAVPSGTGQTSAELEVHPRHSLVSFVVRIVPSPDW
                     FVGIDSLDLCEGGRWKEQVVLDLYPHDAGTDSGFTFSSPNFATIPQDTVTEITASSPS
                     HPANSFYYPRLKSLPPIAKVTFVRLRQSPRAFAPPSLDLASRGNEIVDSLSVPETPLD
                     CEVSLWSSWGLCGGPCGKLGAKSRTRYVRVQPANNGTPCPELEEEAECAPDNCV"
     sig_peptide     512. .577
     misc_feature    578. .1321
                     /note="Region: spondin domain"
     mat_peptide     578. .1501
                     /product="mindin"
     misc_feature    1322. .1501
                     /note="Region: thrombospondin type 1 repeat"
BASE COUNT      553 a    637 c    641 g    494 t
ORIGIN
```

```
  Query Match           60.2%;  Score 665.6;  DB 10;  Length 2325;
  Best Local Similarity  80.1%;  Pred. No. 3.1e-87;
  Matches  795;  Conservative    0;  Mismatches  194;  Indels    3;  Gaps    1;

Qy     44 ccctgggcaaggccctctgcgctctcctcctggccactctcgg---cgccgccggccagc 100
Db    531 CCCTGGACAGAACTCTTTGGGTCTTCCTCCTGGCCATGCTAGGCTCCACCGCAGGCCAGC 590

Qy    101 ctcttgggggagagtccatctgttccgccagagccctggccaaatacagcatcaccttca 160
Db    591 CGCTGGGGGGAGAATCTGTCTGCACAGCCCGGCCTCTGGCTAGATACAGCATCACTTTCA 650

Qy    161 cgggcaagtggagccagacggccttccccaagcagtacccccctgttccgccccccctgccc 220
Db    651 CTGGCAAGTGGAGCCAGACAGCATTTCCCAAGCAGTATCCTCTGTTCCGGCCCCCTGCAC 710

Qy    221 agtggtcttcgctgctgggggccgcgcatagctccgactacagcatgtggaggaagaacc 280
```

```
Db    711 AGTGGTCCTCTCTGCTGGGGGCAGCACACAGCTCTGACTACAGCATGTGGCGGAAGAATG 770

Qy    281 agtacgtcagtaacgggctgcgcgactttgcggagcgcggcgaggcctgggcgctgatga 340
Db    771 AGTACGTCAGCAATGGGCTGAGGGACTTCGCTGAGCGTGGTGAGGCCTGGGCACTGATGA 830

Qy    341 aggagatcgaggcggcgggggaggcgctgcagagcgtgcacgcggtgttttcggcgcccg 400
Db    831 AAGAGATCGAAGCAGCAGGAGAGAAGCTCCAGAGTGTGCACGCGGTATTCTCGGCTCCCG 890

Qy    401 ccgtccccagcggcaccgggcagacgtcggcggagctggaggtgcagcgcaggcactcgc 460
Db    891 CCGTGCCTAGTGGTACCGGGCAGACTTCTGCAGAGTTAGAGGTGCACCCTAGGCACTCGC 950

Qy    461 tggtctcgtttgtggtgcgcatcgtgcccagccccgactggttcgtgggcgtggacagcc 520
Db    951 TGGTGTCCTTCGTGGTACGCATTGTCCCCAGTCCGGACTGGTTTGTGGGCATCGACAGTC 1010

Qy    521 tggacctgtgcgacggggaccgttggcgggaacaggcggcgctggacctgtacccctacg 580
Db   1011 TGGACCTGTGCGAGGGAGGCCGCTGGAAGGAGCAGGTGGTCCTTGATCTTTACCCACACG 1070

Qy    581 acgccgggacggacagcggcttcaccttctcctcccccaacttcgccaccatcccgcagg 640
Db   1071 ATGCAGGGACAGACAGTGGCTTCACCTTCTCCTCCCCCAACTTTGCAACCATCCCACAAG 1130

Qy    641 acacggtgaccgagataacgtcctcctctcccagccacccggccaactccttctactacc 700
Db   1131 ACACAGTGACTGAGATAACCGCCTCATCGCCCAGTCACCCAGCCAACTCATTCTACTACC 1190

Qy    701 cgcggctgaaggccctgcctcccatcgccagggtgacactggtgcggctgcgacagagcc 760
Db   1191 CACGCCTCAAGTCCTTGCCACCCATTGCCAAAGTGACCTTCGTGCGGCTACGGCAGAGTC 1250

Qy    761 ccagggccttcatccctcccgccccagtcctgcccagcagggacaatgagattgtagaca 820
Db   1251 CCAGAGCCTTTGCCCCACCTTCCCTGGACCTGGCCAGCAGAGGAAATGAAATCGTTGACA 1310

Qy    821 gcgcctcagttccagaaacgccgctggactgcgaggtctccctgtggtcgtcctggggac 880
Db   1311 GCCTTTCAGTTCCAGAGACACCGCTGGACTGCGAAGTTTCCCTGTGGTCATCCTGGGGAC 1370

Qy    881 tgtgcggaggccactgtgggaggctcgggaccaagagcaggactcgctacgtccgggtcc 940
Db   1371 TGTGCGGAGGGCCATGTGGAAAGTTGGGAGCCAAGAGCAGAACTCGCTATGTGCGTGTTC 1430

Qy    941 agcccgccaacaacgggagcccctgccccgagctcgaagaagaggctgagtgcgtccctg 1000
Db   1431 AGCCTGCTAACAATGGGACTCCCTGTCCCGAGCTTGAAGAAGAGGCTGAGTGTGCCCCAG 1490

Qy   1001 ataactgcgtctaagaccagagccccgcagcc 1032
Db   1491 ATAACTGTGTCTAAACCCAGACCCCAGTAGCC 1522
```

```
RESULT   5
AB006085
LOCUS       AB006085     2116 bp    mRNA            VRT       10-MAY-2001
DEFINITION  Danio rerio mRNA for MINDIN2, complete cds.
ACCESSION   AB006085
VERSION     AB006085.1  GI:2529222
KEYWORDS    MINDIN2.
SOURCE      Danio rerio embryo cDNA to mRNA.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (sites)
  AUTHORS   Higashijima,S., Nose,A., Eguchi,G., Hotta,Y. and Okamoto,H.
  TITLE     Mindin/F-spondin family: novel ECM proteins expressed in the
            zebrafish embryonic axis
  JOURNAL   Dev. Biol. 192 (2), 211-227 (1997)
  MEDLINE   98104230
REFERENCE   2  (bases 1 to 2116)
  AUTHORS   Higashijima,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-1997) Shin-ichi Higashijima, National Institute
            for Basic Biology, Division of Morphogenesis; Nishigonaka38,
            Myodaijicho, Okazaki, Aichi 444, Japan (E-mail:shinichi@nibb.ac.jp,
            Tel:81-564-55-7572, Fax:81-564-55-7571)
FEATURES             Location/Qualifiers
     source          1. .2116
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /dev_stage="embryo"
     CDS             659. .1654
                     /codon_start=1
                     /product="MINDIN2"
                     /protein_id="BAA22809.1"
                     /db_xref="GI:2529223"
                     /translation="METMTSLRVNCWLTMTLALLSGVPAMPVDVDRMCTAPSTAKYRL
                     TFTGQWTQTAFPKHYPLYRPPAQWSPLIGVTHSSDYHLWQRNEYASNGVREFSERAEA
                     WTLIKEVEAAGERIQSVYGLFSAPAVVAGTGHATTEFEVFARHSLLSFIVRIVPSPDW
                     FIGVDSLNLCEGDHGKENISLELYPYDAGTDSGFTFSSPNFETIPQDKVTQITSSFPS
                     HPANSFYYPRLKHLPPIAKVSLTKIKNNQIFSLPIQPTQSNQIPSGNEIDGPLINTPL
                     DCEVSVWSPWGLCKGQCGEKGVKHRTRYIHMHPANNGAPCPSLEEKRLCIPDNCV"
BASE COUNT      633 a    478 c    413 g    592 t
ORIGIN
```

```
  Query Match             32.8%;  Score 362.6;  DB 5;  Length 2116;
  Best Local Similarity   62.0%;  Pred. No. 1.4e-43;
  Matches  591;  Conservative    0;  Mismatches  359;  Indels    3;  Gaps    1;

Qy     85 ggcgccgccggccagcctcttgggggagagtccatctgttccgccagagccctggccaaa 144
Db    722 GGAGTCCCTGCAATGCCGGTGGATGTGGACCGCATGTGCACGGCCCCTTCCACGGCTAAG 781

Qy    145 tacagcatcaccttcacgggcaagtggagccagacggccttccccaagcagtacccccctg 204
Db    782 TACCGACTGACGTTTACTGGCCAGTGGACACAGACTGCCTTCCCCAAGCACTACCCTCTG 841

Qy    205 ttccgccccccctgcccagtggtcttcgctgctgggggccgcgcatagctccgactacagc 264
Db    842 TACAGGCCACCTGCTCAGTGGTCTCCTCTTATTGGAGTAACTCACAGTTCAGACTATCAC 901

Qy    265 atgtggaggaagaaccagtacgtcagtaacgggctgcgcgactttgcggagcgcggcgag 324
Db    902 CTCTGGCAAAGGAATGAGTACGCCAGTAATGGGGTGAGAGAGTTCTCTGAGAGAGCAGAG 961

Qy    325 gcctgggcgctgatgaaggagatcgaggcggcgggggaggcgctgcagagcgtgcacgcg 384
Db    962 GCTTGGACCCTGATAAAGGAAGTGGAGGCAGCTGGGGAACGGATCCAGAGCGTCTATGGC 1021

Qy    385 gtgttttcggcgcccgccgtccccagcggcaccgggcagacgtcggcggagctggaggtg 444
Db   1022 CTCTTCTCAGCGCCAGCTGTGGTTGCAGGAACTGGCCACGCAACCACAGAATTTGAGGTC 1081

Qy    445 cagcgcaggcactcgctggtctcgtttgtggtgcgcatcgtgcccagccccgactggttc 504
Db   1082 TTTGCGAGACACTCTTTGCTGTCCTTCATTGTTCGCATTGTGCCAAGTCCTGACTGGTTT 1141

Qy    505 gtgggcgtggacagcctggacctgtgcgacggggaccgttggcgggaacaggcggcgctg 564
Db   1142 ATTGGAGTAGACAGTCTGAACTTGTGCGAGGGCGACCACGGTAAAGAAAACATAAGCCTA 1201

Qy    565 gacctgtacccctacgacgccgggacggacagcggcttcaccttctcctcccccaacttc 624
Db   1202 GAGCTCTACCCCTATGATGCCGGCACAGACAGTGGTTTCACCTTCTCCTCGCCAAACTTT 1261

Qy    625 gccaccatcccgcaggacacggtgaccgagataacgtcctcctctcccagccacccggcc 684
Db   1262 GAAACCATTCCTCAAGACAAGGTCACTCAGATAACCTCATCATTCCCAAGCCACCCAGCA 1321

Qy    685 aactccttctactacccgcggctgaaggccctgcctcccatcgccagggtgacactggtg 744
```

```
Db   1322 AACTCTTTCTACTATCCCAGATTGAAGCACCTCCCCCCAATTGCCAAAGTCTCTCTAACC 1381

Qy    745 cggctgcgacagagccccagggccttcatccctcccgccccagtcctgcccagcagggac 804
Db   1382 AAGATCAAGAACAACCAGATCTTCAGCTTACCCATCCAGCCAACACAGTCCAATCAGATC 1441

Qy    805 ---aatgagattgtagacagcgcctcagttccagaaacgccgctggactgcgaggtctcc 861
Db   1442 CCAAGCGGGAATGAAATAGACGGGCCTCTTATAAATACCCCTCTGGACTGCGAGGTGTCT 1501

Qy    862 ctgtggtcgtcctggggactgtgcggaggccactgtgggaggctcgggaccaagagcagg 921
Db   1502 GTGTGGTCACCTTGGGGTCTTTGTAAGGGTCAGTGTGGTGAGAAAGGGGTAAAACACAGA 1561

Qy    922 actcgctacgtccgggtccagcccgccaacaacgggagcccctgccccgagctcgaagaa 981
Db   1562 ACCCGTTATATTCACATGCATCCTGCAAACAACGGTGCTCCCTGTCCATCCCTAGAGGAA 1621

      982 gaggctgagtgcgtccctgataactgcgtctaagaccagagccccgcagcccc 1034
Db   1622 AAGAGACTCTGTATCCCTGACAACTGTGTGTGAATGTAAATCAGATGAGCCTC 1674


RESULT   6
AR037874
LOCUS       AR037874      400 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 1 from patent US 5804382.
ACCESSION   AR037874
VERSION     AR037874.1  GI:5956591
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Sytkowski,A.J. and Yang,M.
  TITLE     Methods for identifying differentially expressed genes and
            differences between genomic nucleic acid sequences
  JOURNAL   Patent: US 5804382-A 1 08-SEP-1998;
FEATURES             Location/Qualifiers
     source          1. .400
                     /organism="unknown"
BASE COUNT       70 a    157 c    110 g     63 t
ORIGIN


  Query Match           31.1%;  Score 343.8;  DB 6;  Length 400;
  Best Local Similarity   97.0%;  Pred. No. 1.6e-40;
  Matches  393;  Conservative    0;  Mismatches    7;  Indels    5;  Gaps    4;

Qy    570 gtacccctacgacgccgggacggacagcggcttcaccttctcctcccccaacttcgccac 629
Db      1 GTACCCCTACGACGCCGGGAC-GACAGCGGCTTCACCTTCTCCTCCCCCAACTTCGCCAC 59

Qy    630 catcccgcaggacacggtgaccgagataacgtcctcctctcccagccacccggccaactc 689
Db     60 CATCCCGCA-GACACGGTGACCGAGATAACGTCCTCCTCTCCCAGCCACCCGGCCAGCTC 118

Qy    690 cttctactacccgcggctgaaggccctgcctcccatcgccagggtgacactggtgcggct 749
Db    119 CTTCTACTACCCGCGGCTGAAGGCC--TGCTCCCATCGCCAGGGTGACACTGGTGCGGCT 176

Qy    750 gcgacagagccccagggccttcatccctcccgccccagtcctgcccagcagggacaatga 809
Db    177 GCGACAGAGCCCCAGGGCCTTCATCCCTCCCGCCCCAGTCCTGCCCAGCA-GGACAATGA 235

Qy    810 gattgtagacagcgcctcagttccagaaacgccgctggactgcgaggtctccctgtggtc 869
Db    236 GATTGTAGACAGCGCCTCAGTTCCAGAAACGCCGCTGGACTGCGAGGTCTCCCTGTGGTC 295

Qy    870 gtcctggggactgtgcggaggccactgtgggaggctcgggaccaagagcaggactcgcta 929
Db    296 GTCCTGGGGACTGTGCGGAGGCCACTGTGGGAGGCTCGGGTCCAAGAGCAGGACTCCGTA 355

Qy    930 cgtccgggtccagcccgccaacaacgggagcccctgccccgagct 974
Db    356 CGCCCGGGTCCAGCCCGCCAACAACGGGAGCCCCTGCCCCGAGCT 400


RESULT   7
AR125230
LOCUS       AR125230      400 bp    DNA             PAT       16-MAY-2001
DEFINITION  Sequence 1 from patent US 6177244.
ACCESSION   AR125230
VERSION     AR125230.1  GI:14111292
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Sytkowski,A.J. and Yang,M.
  TITLE     NPG-1 gene that is differentially expressed in prostate tumors
  JOURNAL   Patent: US 6177244-A 1 23-JAN-2001;
FEATURES             Location/Qualifiers
     source          1. .400
                     /organism="unknown"
BASE COUNT       69 a    159 c    109 g     63 t
ORIGIN


  Query Match           30.7%;  Score 339;  DB 6;  Length 400;
  Best Local Similarity   96.3%;  Pred. No. 8e-40;
  Matches  390;  Conservative    0;  Mismatches   10;  Indels    5;  Gaps    4;

Qy    570 gtacccctacgacgccgggacggacagcggcttcaccttctcctcccccaacttcgccac 629
Db      1 GTACCCCTACGACGCCGGGAC-GACAGCGGCTTCACCTTCTCCTCCCCCAACTTCGCCAC 59

Qy    630 catcccgcaggacacggtgaccgagataacgtcctcctctcccagccacccggccaactc 689
Db     60 CATCCCGCA-GACACGGTGACCGAGATAACGTCCTCCTCTCCCAGCCACCCGGCCAGCTC 118

Qy    690 cttctactacccgcggctgaaggccctgcctcccatcgccagggtgacactggtgcggct 749
Db    119 CTTCTACTACCCGCGGCTGAAGGCC--TGCTCCCATCGCCAGGGTGACACTGGTGCGGCT 176

Qy    750 gcgacagagccccagggccttcatccctcccgccccagtcctgcccagcagggacaatga 809
Db    177 GCGACAGAGCCCCAGGGCCTTCATCCCTCCCGCCCCAGTCCTGCCCAGCA-GGACAATGC 235

Qy    810 gattgtagacagcgcctcagttccagaaacgccgctggactgcgaggtctccctgtggtc 869
Db    236 GCTTGTAGACAGCGCCTCAGTTCCAGAAACACCGCTGGACTGCGAGGTCTCCCTGTGGTC 295

Qy    870 gtcctggggactgtgcggaggccactgtgggaggctcgggaccaagagcaggactcgcta 929
Db    296 GTCCTGGGGACTGTGCGGAGGCCACTGTGGGAGGCTCGGGTCCAAGAGCAGGACTCCGTA 355

Qy    930 cgtccgggtccagcccgccaacaacgggagcccctgccccgagct 974
Db    356 CGCCCGGGTCCAGCCCGCCAACAACGGGAGCCCCTGCCCCGAGCT 400


RESULT   8
AC022763/c
LOCUS       AC022763   197176 bp    DNA             HTG       27-APR-2000
DEFINITION  Homo sapiens clone RP11-296G16, WORKING DRAFT SEQUENCE, 24
            unordered pieces.
ACCESSION   AC022763
VERSION     AC022763.2  GI:7652738
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 197176)
```

```
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-296G16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 197176)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 27, 2000 this sequence version replaced gi:6987625.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5843
                Center clone name: 296_G_16
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 185565 bases at least Q40
                Consensus quality: 190999 bases at least Q30
                Consensus quality: 192979 bases at least Q20
                Insert size: 188000; agarose-fp
                Insert size: 194876; sum-of-contigs
                Quality coverage: 4.5 in Q20 bases; agarose-fp
                Quality coverage: 4.4 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 24 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      438: contig of 438 bp in length
            *      439 538: gap of      100 bp
            *      539     2494: contig of 1956 bp in length
            *     2495 2594: gap of      100 bp
            *     2595     3740: contig of 1146 bp in length
            *     3741 3840: gap of      100 bp
            *     3841     6181: contig of 2341 bp in length
            *     6182 6281: gap of      100 bp
            *     6282     8921: contig of 2640 bp in length
            *     8922 9021: gap of      100 bp
            *     9022    12132: contig of 3111 bp in length
            *    12133 12232: gap of      100 bp
            *    12233    15557: contig of 3325 bp in length
            *    15558 15657: gap of      100 bp
            *    15658    19986: contig of 4329 bp in length
            *    19987 20086: gap of      100 bp
            *    20087    21983: contig of 1897 bp in length
            *    21984 22083: gap of      100 bp
            *    22084    26850: contig of 4767 bp in length
            *    26851 26950: gap of      100 bp
            *    26951    31317: contig of 4367 bp in length
            *    31318 31417: gap of      100 bp
            *    31418    36013: contig of 4596 bp in length
            *    36014 36113: gap of      100 bp
            *    36114    41819: contig of 5706 bp in length
            *    41820 41919: gap of      100 bp
            *    41920    46044: contig of 4125 bp in length
            *    46045 46144: gap of      100 bp
            *    46145    52849: contig of 6705 bp in length
            *    52850 52949: gap of      100 bp
            *    52950    63466: contig of 10517 bp in length
            *    63467 63566: gap of      100 bp
            *    63567    70466: contig of 6900 bp in length
            *    70467 70566: gap of      100 bp
            *    70567    80411: contig of 9845 bp in length
            *    80412 80511: gap of      100 bp
            *    80512    92230: contig of 11719 bp in length
            *    92231 92330: gap of      100 bp
            *    92331   105063: contig of 12733 bp in length
            *   105064 105163: gap of      100 bp
            *   105164   115199: contig of 10036 bp in length
            *   115200 115299: gap of      100 bp
            *   115300   127901: contig of 12602 bp in length
            *   127902 128001: gap of      100 bp
            *   128002   145911: contig of 17910 bp in length
            *   145912 146011: gap of      100 bp
            *   146012   197176: contig of 51165 bp in length.
FEATURES             Location/Qualifiers
     source          1. .197176
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-296G16"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .438
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    539. .2494
                     /note="assembly_fragment"
     misc_feature    2595. .3740
                     /note="assembly_fragment"
     misc_feature    3841. .6181
                     /note="assembly_fragment"
     misc_feature    6282. .8921
                     /note="assembly_fragment"
     misc_feature    9022. .12132
                     /note="assembly_fragment"
     misc_feature    12233. .15557
                     /note="assembly_fragment"
     misc_feature    15658. .19986
                     /note="assembly_fragment"
     misc_feature    20087. .21983
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    22084. .26850
                     /note="assembly_fragment"
     misc_feature    26951. .31317
                     /note="assembly_fragment"
     misc_feature    31418. .36013
                     /note="assembly_fragment"
     misc_feature    36114. .41819
                     /note="assembly_fragment"
     misc_feature    41920. .46044
                     /note="assembly_fragment"
     misc_feature    46145. .52849
                     /note="assembly_fragment"
     misc_feature    52950. .63466
                     /note="assembly_fragment"
     misc_feature    63567. .70466
```

```
                     /note="assembly_fragment"
     misc_feature    70567. .80411
                     /note="assembly_fragment"
     misc_feature    80512. .92230
                     /note="assembly_fragment"
     misc_feature    92331. .105063
                     /note="assembly_fragment"
     misc_feature    105164. .115199
                     /note="assembly_fragment"
     misc_feature    115300. .127901
                     /note="assembly_fragment"
     misc_feature    128002. .145911
                     /note="assembly_fragment"
     misc_feature    146012. .197176
                     /note="assembly_fragment"
BASE COUNT    46883 a  49476 c  49808 g  48706 t   2303 others
ORIGIN


  Query Match             30.4%;  Score 336.2;  DB 2;  Length 197176;
  Best Local Similarity   82.7%;  Pred. No. 1e-40;
  Matches  429;  Conservative    0;  Mismatches   18;  Indels   72;  Gaps    1;

Qy     212 cccctgcccagtggtcttcgctgctgggggccgcgcatagctccgactacagcatgtgga 271
           || | || | | |  | |   |     ||||||||||||||||||||||||||||||||
Db  133359 CCGCGGCCGCGGCCGGGTGGTCCTGCAGGGGCCGCGCATAGCTCCGACTACAGCATGTGGA 133300

Qy     272 ggaagaaccagtacgtcagtaacgggctgcgcgactttgcggagcgcggcgaggcctggg 331
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133299 GGAAGAACCAGTACGTCAGTAACGGGCTGCGCGACTTTGCGGAGCGCGGCGAGGCCTGGG 133240

Qy     332 cgctgatgaaggagatcgaggcggcgggggaggcgctgcagagcgtgcacgcggtgtttt 391
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133239 CGCTGATGAAGGAGATCGAGGCGGCGGGGGAGGCGCTGCAGAGCGTGCACGCGGTGTTTT 133180

Qy     392 cggcgcccgccgtccccagcggcaccgggcagacgtcggcggagctggaggtgcagcgca 451
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133179 CGGCGCCCGCCGTCCCCAGCGGCACCGGGCAGACGTCGGCGGAGCTGGAGGTGCAGCGCA 133120

Qy     452 ggcactcgct---------------------------------------------------- 461
           ||||||||||
Db  133119 GGCACTCGCTGGTGAGCGCGGGGCCGGCCGGGGGCGGGGTCCGGGGGGCCGGGGCCAGGC 133060

Qy     462 ----------------------ggtctcgtttgtggtgcgcatcgtgcccagccccgact 499
                                  ||||||||||||||||||||||||||||||||||||
Db  133059 GCGCTGAGCCGCTGTCGCCGCAGGTCTCGTTTGTGGTGCGCATCGTGCCCAGCCCCGACT 133000

Qy     500 ggttcgtgggcgtggacagcctggacctgtgcgacggggaccgttggcgggaacaggcgg 559
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132999 GGTTCGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGGACCGTTGGCGGGAACAGGCGG 132940

Qy     560 cgctggacctgtacccctacgacgccgggacggacagcggcttcaccttctcctccccca 619
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132939 CGCTGGACCTGTACCCCTACGACGCCGGGACGGACAGCGGCTTCACCTTCTCCTCCCCCA 132880

Qy     620 acttcgccaccatcccgcaggacacggtgaccgagataa 658
           |||||||||||||||||||||||||||||||||| | |
Db  132879 ACTTCGCCACCATCCCGCAGGACACGGTGACCGAGGTGA 132841


RESULT   9
AR035966
LOCUS       AR035966      506 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 15 from patent US 5871969.
ACCESSION   AR035966
VERSION     AR035966.1  GI:5952634
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 506)
  AUTHORS   Hastings,G. and Dillon,P.J.
  TITLE     Nucleic acids encoding human neuronal attachment factor-1
  JOURNAL   Patent: US 5871969-A 15 16-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .506
                     /organism="unknown"
BASE COUNT       91 a    136 c    150 g    100 t     29 others
ORIGIN


  Query Match             27.1%;  Score 300;  DB 6;  Length 506;
  Best Local Similarity   84.8%;  Pred. No. 3e-34;
  Matches  368;  Conservative    0;  Mismatches   62;  Indels    4;  Gaps    4;

Qy      13 cgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctctgcgctctcctc 72
           |||   | |  ||||||||||||||| ||||||||||||||||||  ||||||||||||
Db       6 CGGCANAGGNNAAACCCCAGCCCGGCTGCCGCCCTGGGCAAGGCCTNCTGCGCTCTCCTC 65

Qy      73 ctggccactctcggcgccgccggccagcctcttgggggagagtccatctgttccgccaga 132
           |||||||||||||||||| |  |||||||||||||||||||||||||| ||||||||||
Db      66 CTGGCCACTCTCGGCGCCGGC-ACCAGCCTCTTGGGGGAGAGTCCATCTNTTCCGCCAGA 124

Qy     133 gccctggccaaatacagcatcaccttcacgggcaagtggagccagacggccttccccaag 192
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     125 GCCCCGGCCAAATACAGCATCACCTTCACGGGCAAGTGGAGCCAGACGGCCTTCCCCAAG 184

Qy     193 cagtacccccctgttccgccccccctgcccagtggtcttcgctgctgggggccgcgcatagc 252
           |||||||||||||||||||||||||| || |||| ||||||||||||||||||||||||
Db     185 CAGTACCCCCTGTTCCGCCCCCCCTGCGCA-TGGTNTTCGCTGCTGGGGGCCGCGCATAGC 243

Qy     253 tccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctgcgcgactttgcg 312
           ||||||||||||||||||||||||||||||||||  ||||||||||||||||| ||
Db     244 TCCGACTACAGCATGTGGAGGAAGAACCAGTACGTCATAAACGGGCTGCGCGACTTTNCG 303

Qy     313 gagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggggaggcgctgcag 372
           ||||||  | | | |  || || ||||||||||||||  || ||||||||||||||    |
Db     304 GAGCGGCGAG-GCCTNGGNCGTTGATGAAGGAGATCCGGGNGGCGGGGGAGGCGTNCAAN 362

Qy     373 agcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccgggcagacgtcggcg 432
           ||    || | | | |||||||| |||||   ||||||  || | | || | |||| || |
Db     363 AGGTGNCAAGAGTTNTTTTCGGGGCCCG-GTTCCCCAANGGNAACNGGNAAACGTTGGGG 421

Qy     433 gagctggaggtgca 446
           |   |  || |  |
Db     422 GNTTTNNAGTTTNA 435


RESULT  10
AK024499
LOCUS       AK024499     4292 bp    mRNA            PRI       29-SEP-2000
DEFINITION  Homo sapiens mRNA for FLJ00108 protein, partial cds.
ACCESSION   AK024499
VERSION     AK024499.1  GI:10440517
KEYWORDS    fis (full insert sequence).
SOURCE      Homo sapiens adult spleen cDNA to mRNA, clone:as00108.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4292)
  AUTHORS   Ohara,O., Nagase,T., Kikuno,R. and Okumura,K.
  TITLE     The nucleotide sequence of a long cDNA clone isolated from human
            spleen
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 4292)
  AUTHORS   Ohara,O., Nagase,T., Kikuno,R. and Okumura,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2000) to the DDBJ/EMBL/GenBank databases. Osamu
            Ohara, Kazusa DNA Research Institute, Department of Human Gene
            Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:cdnainfo@kazusa.or.jp, URL:http://www.kazusa.or.jp/NEDO,
            Tel:81-438-52-3913, Fax:81-438-52-3914)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
```

```
          Kazusa DNA Research Institute.
FEATURES             Location/Qualifiers
     source          1. .4292
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="as00108"
                     /dev_stage="adult"
                     /tissue_type="spleen"
                     /note="vector:pBluescriptII SK plus"
     gene            1. .653
                     /gene="FLJ00108"
     CDS             <1. .653
                     /gene="FLJ00108"
                     /note="For this clone, GeneMark analysis triggered an
                     alert for spurious CDS split. The result of GeneMark
                     analysis of the nucleotide sequence of this clone is
                     accessible through http://www.kazusa.or.jp/NEDO.
                     Start codon is not identified."
                     /codon_start=3
                     /product="FLJ00108 protein"
                     /protein_id="BAB15789.1"
                     /db_xref="GI:10440518"
                     /translation="SAELEVQRRHSLVSFVVRIVPSPDWFVGVDSLDLCDGDRWREQA
                     ALDLYPYDAGTDSGFTFSSPNFATIPQDTVTEVSGHPACARAAMAPAVGTGLGPGCTL
                     ASQPRTAACGFCVEEALLLAPAVSKGVQPSHTLPTRGHSTAGRTAGRASRGGRPFLAS
                     VSSPCSVGWLRGVHGTLSPTVVQRVRWTLGSGPSEA1RIPEGGLLGPQANLFFLPR"
BASE COUNT      788 a   1323 c   1349 g    832 t
ORIGIN
```

```
  Query Match             25.2%;  Score 279;  DB 9;  Length 4292;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-31;
  Matches  279;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     827 cagttccagaaacgccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcg 886
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3568 CAGTTCCAGAAACGCCGCTGGACTGCGAGGTCTCCCTGTGGTCGTCCTGGGGACTGTGCG 3627

Qy     887 gaggccactgtgggaggctcgggaccaagagcaggactcgctacgtccgggtccagcccg 946
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3628 GAGGCCACTGTGGGAGGCTCGGGACCAAGAGCAGGACTCGCTACGTCCGGGTCCAGCCCG 3687

Qy     947 ccaacaacgggagcccctgccccgagctcgaagaagaggctgagtgcgtccctgataact 1006
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3688 CCAACAACGGGAGCCCCTGCCCCGAGCTCGAAGAAGAGGCTGAGTGCGTCCCTGATAACT 3747

Qy    1007 gcgtctaagaccagagccccgcagcccctggggccccccggagccatggggtgtcggggg 1066
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3748 GCGTCTAAGACCAGAGCCCCGCAGCCCCTGGGGCCCCCCGGAGCCATGGGGTGTCGGGGG 3807

Qy    1067 ctcctgtgcaggctcatgctgcaggcggccgagggcaca 1105
           |||||||||||||||||||||||||||||||||||||||
Db    3808 CTCCTGTGCAGGCTCATGCTGCAGGCGGCCGAGGGCACA 3846
```

```
RESULT  11
AK026054
LOCUS       AK026054     2820 bp    mRNA            PRI       29-SEP-2000
DEFINITION  Homo sapiens cDNA: FLJ22401 fis, clone HRC08032, highly similar to
            AB027466 Homo sapiens SPON2 mRNA for spondin 2.
ACCESSION   AK026054
VERSION     AK026054.1  GI:10438772
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens primary human renal epithelial cells cDNA to mRNA,
            clone_lib:HRC clone:HRC08032.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
            Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
            Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 2820)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
            Sugano, Institute of Medical Science, University of Tokyo,
            Laboratory of Genome Structure Analysis, Human Genome Center;
            Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
            (E-mail:cdna1@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Departent of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1. .2820
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="primary human renal epithelial cells"
                     /clone="HRC08032"
                     /clone_lib="HRC"
                     /note="cloning vector pME18SFL3"
     misc_feature    1. .2820
                     /note="highly similar to AB027466 Homo sapiens SPON2 mRNA
                     for spondin 2"
BASE COUNT      535 a    849 c    873 g    563 t
ORIGIN
```

```
  Query Match             25.1%;  Score 277.4;  DB 9;  Length 2820;
  Best Local Similarity   99.6%;  Pred. No. 2.3e-31;
  Matches  278;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     827 cagttccagaaacgccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcg 886
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2079 CAGTTCCAGAAACGCCGCTGGACTGCGAGGTCTCCCTGTGGTCGTCCTGGGGACTGTGCG 2138

Qy     887 gaggccactgtgggaggctcgggaccaagagcaggactcgctacgtccgggtccagcccg 946
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2139 GAGGCCACTGTGGGAGGCTCGGGACCAAGAGCAGGACTCGCTACGTCCGGGTCCAGCCCG 2198

Qy     947 ccaacaacgggagcccctgccccgagctcgaagaagaggctgagtgcgtccctgataact 1006
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2199 CCAACAACGGGAGCCCCTGCCCCGAGCTCGAAGAAGAGGCTGAGTGCGTCCCTGATAACT 2258

Qy    1007 gcgtctaagaccagagccccgcagcccctggggccccccggagccatggggtgtcggggg 1066
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2259 GCGTCTAAGACCAGAGCCCCGCAGCCCCTGGGGCCCCCCGGAGCCATGGGGTGTCGGGGG 2318

Qy    1067 ctcctgtgcaggctcatgctgcaggcggccgagggcaca 1105
           |||||||||||||||||| ||||||||||||||||||||
Db    2319 CTCCTGTGCAGGCTCATACTGCAGGCGGCCGAGGGCACA 2357
```

```
RESULT  12
AB006084
LOCUS       AB006084     3381 bp    mRNA            VRT       10-MAY-2001
DEFINITION  Danio rerio mRNA for MINDIN1, complete cds.
ACCESSION   AB006084
VERSION     AB006084.1  GI:2529220
KEYWORDS    MINDIN1.
SOURCE      Danio rerio embryo cDNA to mRNA.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (sites)
  AUTHORS   Higashijima,S., Nose,A., Eguchi,G., Hotta,Y. and Okamoto,H.
  TITLE     Mindin/F-spondin family: novel ECM proteins expressed in the
```

```
            zebrafish embryonic axis
  JOURNAL   Dev. Biol. 192 (2), 211-227 (1997)
  MEDLINE   98104230
REFERENCE   2  (bases 1 to 3381)
  AUTHORS   Higashijima,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-1997) Shin-ichi Higashijima, National Institute
            for Basic Biology, Division of Morphogenesis; Nishigonaka38,
            Myodaijicho, Okazaki, Aichi 444, Japan (E-mail:shinichi@nibb.ac.jp,
            Tel:81-564-55-7572, Fax:81-564-55-7571)
FEATURES             Location/Qualifiers
     source          1. .3381
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /dev_stage="embryo"
     CDS             48. .1052
                     /codon_start=1
                     /product="MINDIN1"
                     /protein_id="BAA22808.1"
                     /db_xref="GI:2529221"
                     /translation="MMSSEILVPGWLQQLLVVLLRFTLSCAALVNSTNGTECSARGPA
                     SYIVVFTGHWSPQTFPKQYPLFRPPAQWSKLMVVTHNEQYRLWQEGAPASDGMKSFAE
                     QGLTVDLVKDAKEARKRRSVGSMYRTAGIPSGIGHSSTEVLLTPRSPLVSLIVKLIPS
                     PDWFVGVDGLNLCEGGKWKQEVTFDLHPFDAGTDSGFTFSSPNFPTTPPENITMITSQ
                     KPNHPANSFYYPRLNELPPLATIWVKRQSRLPVRQQNRLSNHILPDASKPHRFSETPL
                     DCEVSMWSSWGLCFGPCARGGLRHRTRYILLKPANSGSPCPELEEQEECTPHNCLADQ
                     "
     polyA_site      3381
                     /note="11 a nucleotides"
BASE COUNT      949 a    713 c    707 g   1012 t
ORIGIN
```

```
  Query Match             23.3%;  Score 258;  DB 5;  Length 3381;
  Best Local Similarity   57.1%;  Pred. No. 1.3e-28;
  Matches  512;  Conservative    0;  Mismatches  375;  Indels    9;  Gaps    2;

Qy    127 gccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacggccttc 186
          ||  |||  |  ||    |||| | |    ||||| ||  | ||||| |    | |||||
Db    165 GCTCGAGGACCTGCATCTTACATCGTGGTGTTCACCGGACACTGGAGTCCACAGACCTTC 224

Qy    187 cccaagcagtacccccctgttccgccccccctgcccagtggtcttcgctgctgggggccgcg 246
          || || ||||| ||| | ||||| || ||||| |||||||||    ||  |||  |   |
Db    225 CCTAAACAGTATCCCTTATTCCGGCCACCTGCGCAGTGGTCCAAACTCATGGTTGTAACC 284

Qy    247 catagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctgcgcgac 306
          ||||      | ||| |  ||||| | ||     |   |  ||  | ||  ||     |
Db    285 CATAATGAGCAGTACCGGCTGTGGCAGGAGGGGGCGCCGGCGAGCGATGGCATGAAGAGC 344

Qy    307 tttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggggaggcg 366
          || || ||||  ||     |   ||  ||| |||||||     ||| |||| || ||  |
Db    345 TTCGCAGAGCAGGGGCTCACAGTGGATCTGGTGAAGGATGCGAAGGAGGCGAGGAAGAGG 404

Qy    367 ctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccgggcagacg 426
          |    |   |||    |  |||       || | |   ||||    |||| ||| || |
Db    405 CGATCG---GTGGGTTCCATGTACCGTACGGCTGGAATCCCGTCTGGCATCGGCCACAGT 461

Qy    427 tcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgcatcgtg 486
          ||  | ||| ||    ||   | ||||    | ||||| ||  |  | |||    | |
Db    462 TCCACAGAGGTGCTTCTGACCCCCAGGAGTCCACTGGTGTCTCTGATAGTGAAACTGATC 521

Qy    487 cccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggaccgttgg 546
          ||||| || |||||||| ||||| || ||  | ||  ||||||| || || |     |||
Db    522 CCCAGTCCTGACTGGTTTGTGGGGGTTGATGGTCTAAACCTGTGTGAAGGAGGGAAGTGG 581

Qy    547 cgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggcttcacc 606
          |  | | | |  | ||| || |||| |  || || || || |||||||| |||||
Db    582 AAACAGGAAGTGACCTTTGACTTGCACCCATTTGATGCGGGCACAGACAGCGGATTCACT 641

Qy    607 ttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacgtcctcc 666
          |||||||| || |||||| | || |  || | ||| |   | |||  ||| || ||
Db    642 TTCTCCTCTCCAAACTTCCCAACAACACCTCCGGAAAACATCACCATGATCACCTCTCAA 701
```

```
Qy    667 tctcccagccacccggccaactccttctactacccgcggctgaaggccctgcctcccatc 726
          || | |||||||||| || ||||| || || || ||||| || |  || || ||  |
Db    702 AAGCCAAACCACCCGGCAAATTCCTTTTATTATCCTCGGCTAAATGAACTCCCACCTCT- 760

Qy    727 gccagggtgacactggtgcggctgcgacagagccccagggccttcatccctcccgcccca 786
          ||  ||  |   |  |   ||||||  || | | |  ||   |  |    ||
Db    761 -----GGCAACCATATGGGTGAAACGACAGTCCCGCTTGCCTGTCCGTCAACAGAACCGG 815

Qy    787 gtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacgccgctg 846
          |       ||       ||       |      |    || ||||||||||||||
Db    816 CTGTCCAATCACATCCTACCAGACGCAAGCAAACCTCACAGGTTTTCAGAAACGCCGCTG 875

Qy    847 gactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtgggaggctc 906
          |||||||||| ||  ||||||||||||||| |||||    ||||  || |   |   |
Db    876 GACTGCGAGGTGTCGATGTGGTCGTCCTGGGGTCTGTGTTTTGGCCCGTGCGCTCGAGGC 935

Qy    907 gggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagcccctgc 966
          ||           | ||||| ||| ||| | |  | || ||||||| |||    || ||
Db    936 GGCCTACGCCATCGCACTCGTTACATCCTGCTGAAACCTGCCAACAGCGGTTCGCCATGT 995

Qy    967 cccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccagag 1022
          |||||||| || ||| |||  ||||||   ||  | |||||| |   ||| ||| |
Db    996 CCCGAGCTGGAGGAACAGGAGGAGTGCACACCGCACAACTGCCTGGCAGATCAGTG 1051
```

```
RESULT  13
AR035967
LOCUS       AR035967      316 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 16 from patent US 5871969.
ACCESSION   AR035967
VERSION     AR035967.1  GI:5952635
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 316)
  AUTHORS   Hastings,G. and Dillon,P.J.
  TITLE     Nucleic acids encoding human neuronal attachment factor-1
  JOURNAL   Patent: US 5871969-A 16 16-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .316
                     /organism="unknown"
BASE COUNT       48 a     84 c    120 g     49 t     15 others
ORIGIN
```

```
  Query Match             18.3%;  Score 202.6;  DB 6;  Length 316;
  Best Local Similarity   85.6%;  Pred. No. 4e-20;
  Matches  274;  Conservative    0;  Mismatches   39;  Indels    7;  Gaps    5;

Qy    272 ggaagaaccagtacgtcagtaacgggctgcgcgactttgcggagcgcggcgaggcctggg 331
          || |   |||||||||||| |||||||||||||||||||||||||  ||||||||||||||
Db      1 GGCANNGCCAGTACGTCA-TAACGGGCTGCGCGACTTTGCGGA-NGCGGCGAGGCCTGGG 58

Qy    332 cgctgatgaaggagatcgaggcggcgggggaggcgctgcagagcgtgcacgcggtgtttt 391
          |||||||||||||||||| |||||||||||||||||||||||||| |||||||  ||||||||
Db     59 CGCTGATGAAGGAGATCAAGGCGGCGGGGGAGGCGCTGCAGAG-GTGCACGAGGTGTTTT 117

Qy    392 cggcgcccgccgtccccagcggcaccgggcagacgtcggcggagctggaggtgcagcgca 451
          |||||||||     |||||||| |||| ||||||||||||||| |   | ||||||||||||
Db    118 CGGCGCCCG-GTNNCCCAGCGNCACCNGGCAGACGTCGGCGAACTGGNAGGTGCAGCGCA 176

Qy    452 ggcactcgctggtctcgtttgtggtgcgcatcgtgcccagccccgactggttcgtgggcg 511
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177 GGCACTCGCTGGTCTCGTTTGTGGTGCGCATCGTGCCCAGCCCCGACTGGTTCGTGGGCG 236

Qy    512 tggacagcctggacc---tgtgcgacggggaccgttggcgggaacaggcggcgctggacc 568
          ||||||||||||  |   ||     |||  |  | |   | ||||||| || ||||||
Db    237 TGGACAGCCTGGGACCTGTGANAACGGGGACCTTTNGCGNGNAACAGGCGNCGTTGGACC 296

Qy    569 tgtacccctacgacgccggg 588
          |||| |||||||||| ||||
Db    297 TGTANCCCTACGACGNCGGG 316
```

```
RESULT  14
AR035968
LOCUS       AR035968      316 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 17 from patent US 5871969.
ACCESSION   AR035968
VERSION     AR035968.1  GI:5952636
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 316)
  AUTHORS   Hastings,G. and Dillon,P.J.
  TITLE     Nucleic acids encoding human neuronal attachment factor-1
  JOURNAL   Patent: US 5871969-A 17 16-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .316
                     /organism="unknown"
BASE COUNT       48 a     84 c    120 g     49 t     15 others
ORIGIN

  Query Match             18.3%;  Score 202.6;  DB 6;  Length 316;
  Best Local Similarity   85.6%;  Pred. No. 4e-20;
  Matches  274;  Conservative    0;  Mismatches   39;  Indels    7;  Gaps    5;

Qy     272 ggaagaaccagtacgtcagtaacgggctgcgcgactttgcggagcgcggcgaggcctggg 331
           || |   |||||||||||| |||||||||||||||||||||||||  ||||||||||||||||
Db       1 GGCANNGCCAGTACGTCA-TAACGGGCTGCGCGACTTTGCGGA-NGCGGCGAGGCCTGGG 58

Qy     332 cgctgatgaaggagatcgaggcggcgggggaggcgctgcagagcgtgcacgcggtgtttt 391
           |||||||||||||||||| |||||||||||||||||||||||| ||||||| ||||||||
Db      59 CGCTGATGAAGGAGATCAAGGCGGCGGGGGAGGCGCTGCAGAG-GTGCACGAGGTGTTTT 117

Qy     392 cggcgcccgccgtccccagcggcaccgggcagacgtcggcggagctggaggtgcagcgca 451
           |||||||||     ||||||| |||| ||||||||||||| |    | ||||||||||||
Db     118 CGGCGCCCG-GTNNCCCAGCGNCACCNGGCAGACGTCGGCGAACTGGNAGGTGCAGCGCA 176

Qy     452 ggcactcgctggtctcgtttgtggtgcgcatcgtgcccagccccgactggttcgtgggcg 511
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     177 GGCACTCGCTGGTCTCGTTTGTGGTGCGCATCGTGCCCAGCCCCGACTGGTTCGTGGGCG 236

Qy     512 tggacagcctggacc---tgtgcgacggggaccgttggcgggaacaggcggcgctggacc 568
           |||||||||||||  |   ||      |||  |  | |   | |||||||| || ||||||
Db     237 TGGACAGCCTGGGACCTGTGANAACGGGGACCTTTNGCGNGNAACAGGCGNCGTTGGACC 296

Qy     569 tgtacccctacgacgccggg 588
           |||| ||||||||||| ||||
Db     297 TGTANCCCTACGACGNCGGG 316

RESULT  15
DMAF000178
LOCUS       DMAF000178   2565 bp    mRNA            INV       08-MAY-2001
DEFINITION  Drosophila melanogaster M-spondin (mspo) mRNA, complete cds.
ACCESSION   AF000178
VERSION     AF000178.1  GI:2130612
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 2565)
  AUTHORS   Umemiya,T., Takeichi,M. and Nose,A.
  TITLE     M-spondin, a novel ECM protein highly homologous to vertebrate
            F-spondin, is localized at the muscle attachment sites in the
            Drosophila embryo
  JOURNAL   Dev. Biol. 186 (2), 165-176 (1997)
  MEDLINE   97349133
   PUBMED   9205137
REFERENCE   2  (bases 1 to 2565)
  AUTHORS   Umemiya,T., Takeichi,M. and Nose,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-1997) Behavior and Neurobiology, National
            Institute for Basic Biology, Myodaiji-cho, Okazaki, Aichi Pref.
            444, Japan
FEATURES             Location/Qualifiers
     source          1. .2565
                     /organism="Drosophila melanogaster"
                     /strain="Canton S"
                     /db_xref="taxon:7227"
                     /chromosome="2"
                     /map="51C; polytene"
                     /dev_stage="9-12 hr or 13-16 hr AEL embryos"
     gene            1. .2565
                     /gene="mspo"
     CDS             346. .2142
                     /gene="mspo"
                     /note="F-spondin-related ECM protein; secreted protein
                     expressed by somatic body wall muscles; localized at the
                     muscle attachment site in late embryos"
                     /codon_start=1
                     /product="M-spondin"
                     /protein_id="AAB58341.1"
                     /db_xref="GI:2130613"
                     /translation="MGATPKTDFPAHWLLCILLALGSLSSKANAVSAYGASFISNPSI
                     YVYSVENQPDERESRDNNIKFAAEHGNLNSNSNLNRNYYYTDDPSPVPNMNPTASLAT
                     PPATQTAPQTGCTLDRLAVYKVVLHTYWTRELFPKHYPDWRPTAQWTKTLGRTHNANY
                     ALYHIGQPATAAVKQFAESGRTDLLDSNAGEQQQVQMQLQSQMQAGKSPSGGISSGTT
                     SFNATAASTATPTGGSGGSGGSGVTGTTTAERSVFDEFSMPAIPMGAGRSEAQVFVDS
                     NHSLVSLMTRIVPSPDWFIGVDSFELCVGGSWIDTVTVELDPLDAGTDNGFTFTAPNW
                     PTAPQGVIYRITSRYPGHPAGSFYYPKSKRLPPIATFQFIKLKEYELSEVFNIAEDDR
                     KYETVQTQTHLDAEHNHVEMNNELSASIERERQTEQQQLQQNDDERQRIRSQLLAKMN
                     PIYGSNNSLQAAPGQVVSVVPKNDKHAILQSIASSYRRAADASDANASKPTPSAIGGG
                     KAGGAVGGGAATRRRSSAQRRRDCRVSHWSEWTACSKSCGVGEMHRYRKVIKHGKRGG
                     RQCPALQQSKWCGTERNCHGSQTYFNWSDSDT"
BASE COUNT      689 a    679 c    677 g    520 t
ORIGIN

  Query Match             10.7%;  Score 118.2;  DB 3;  Length 2565;
  Best Local Similarity   56.6%;  Pred. No. 2.1e-08;
  Matches  219;  Conservative    0;  Mismatches  168;  Indels    0;  Gaps    0;

Qy     364 gcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccgggcag 423
           ||   ||  ||||||  ||  | |||       ||| ||  | || |  ||| |||| |
Db    1087 GCCGAGCGCAGCGTGTTCGATGAGTTCAGCATGCCGGCGATTCCGATGGGCGCCGGACGC 1146

Qy     424 acgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgcatc 483
           |    |||  |  ||   ||  |  |||  |||   |||||||| |  ||   |||||
Db    1147 AGTGAGGCACAAGTGTTTGTCGACAGCAATCACAGTCTGGTCTCGCTAATGACACGCATT 1206

Qy     484 gtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggaccgt 543
           || |||   || |||||||| | ||||||||     | || || ||||  || | |
Db    1207 GTTCCCTCGCCGGACTGGTTCATCGGCGTGGATTCGTTCGAGCTCTGCGTGGGTGGCTCC 1266

Qy     544 tggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggcttc 603
           |||   ||   || | |  | || ||| | || |  || || || ||||| |  || |||
Db    1267 TGGATCGATACGGTGACCGTTGAGCTGGATCCTTTGGATGCGGGCACGGATAATGGGTTC 1326

Qy     604 accttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacgtcc 663
           |||||| |  | |||||||   | ||    || |||| |    |      ||| || ||
Db    1327 ACCTTCACGGCGCCCAACTGGCCAACAGCACCTCAGGGCGTCATCTATCGGATCACCTCG 1386

Qy     664 tcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcctccc 723
           | |||  | || || ||      ||| || || ||| |   ||     |||| |||
Db    1387 CGTTATCCTGGACATCCTGCGGGTAGCTTTTATTATCCGAAGAGCAAACGATTGCCGCCC 1446

Qy     724 atcgccagggtgacactggtgcggctg 750
           ||||||| | |     |  |  ||||
Db    1447 ATCGCCACGTTCCAGTTTATCAAGCTG 1473
```

Search completed: December 10, 2001, 19:46:49
Job time: 22163 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2001, 13:37:26 ; Search time 164.06 Seconds
(without alignments)
5774.378 Million cell updates/sec

Title: US-09-170-042A-1
Perfect score: 1105
Sequence: 1 cgctgctcctgccgggtgat.....tgcaggcggccgagggcaca 1105

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1105	100.0	1105	20	AAX02019	Human NAF-1 DNA.
	2	1100.2	99.6	1105	18	AAT78360	Human neuronal att
	3	1100.2	99.6	1840	21	AAZ95034	Cancer specific ge
	4	1100	99.5	1848	22	AAH34981	Human colon cancer
	5	1098.6	99.4	1607	19	AAV63241	cDNA encoding an a
	6	1098.6	99.4	1770	22	AAF90567	Human extracellula
	7	1098.6	99.4	1785	22	AAF90566	Human extracellula
	8	1098.6	99.4	1846	20	AAZ34089	Human PRO866 nucle
	9	1098.6	99.4	1847	21	AAC78507	Human PRO866 (UNQ4
	10	1098.6	99.4	1847	21	AAC58630	Human PRO866 prote
	11	1098.6	99.4	1847	21	AAA49728	Human PRO866 cDNA
C	12	1098.6	99.4	1926	22	AAI59814	Human polynucleoti
	13	1098.6	99.4	1991	22	AAI58028	Human polynucleoti
	14	1088.2	98.5	1718	21	AAC59794	Human secreted pro
	15	989.6	89.6	996	20	AAV72537	Human mindin cDNA.
	16	843.8	76.4	1458	21	AAZ52575	Human secreted pro
	17	681	61.6	993	19	AAV63258	Degnerate sequence
	18	343.8	31.1	400	19	AAV53726	Nucleotide sequenc
	19	341.6	30.9	1021	20	AAV72538	Human mindin-relat
	20	339	30.7	400	20	AAZ06948	Human NPG-1 partia
	21	300	27.1	506	20	AAX02024	Human NAF-1 DNA ho
	22	279	25.2	2261	21	AAC59258	Human secreted pro
	23	233.4	21.1	432	19	AAV63270	EST sequence encod
	24	204	18.5	204	21	AAZ95016	Prostate cancer sp
	25	202.6	18.3	316	20	AAX02026	Human NAF-1 DNA ho
	26	202.6	18.3	316	20	AAX02025	Human NAF-1 DNA ho
	27	198.4	18.0	553	19	AAV63268	EST sequence encod
	28	176.8	16.0	539	22	AAI57550	Human colorectal c
	29	173.2	15.7	541	19	AAV63266	EST sequence encod
	30	142.2	12.9	470	19	AAV63267	EST sequence encod
	31	78	7.1	4061	20	AAX30094	Bovine vascular sm
	32	71.6	6.5	3079	20	AAX30093	Human vascular smo
	33	71.6	6.5	3079	22	AAF94993	Human ovarian canc
	34	71.6	6.5	3999	22	AAF94986	Human ovarian canc
	35	70	6.3	4029	14	AAQ52674	F-spondin (FP5-9)
	36	70	6.3	4029	20	AAX30095	Rat vascular smoot
	37	61.6	5.6	3226	14	AAQ52675	F-spondin coding s
C	38	61.2	5.5	114955	20	AAX53491	Human adenosine A1
	39	57.2	5.2	1816	14	AAQ52676	Partial sequence o
C	40	57.2	5.2	1848	22	AAH34981	Human colon cancer
C	41	57	5.2	1105	18	AAT78360	Human neuronal att
C	42	57	5.2	1105	20	AAX02019	Human NAF-1 DNA.
C	43	57	5.2	1458	21	AAZ52575	Human secreted pro
C	44	57	5.2	1718	21	AAC59794	Human secreted pro
	45	57	5.2	1926	22	AAI59814	Human polynucleoti

ALIGNMENTS

RESULT 1
AAX02019
ID AAX02019 standard; DNA; 1105 BP.
XX
AC AAX02019;
XX
DT 21-APR-1999 (first entry)
XX
DE Human NAF-1 DNA.
XX
KW NAF-1; neuronal attachment factor-1; F-spondin analogue; treatment;
KW spinal cord injury; peripheral nerves damage; neural cell adhesion;
KW neurite extension; tumour cell metastasis; inhibitor; mobility; disease;
KW endothelial cell proliferation; tumour neovascularisation; haemostasis;
KW angiostatic agent; wound healing; diagnostic; neurotrophic; anticancer;
KW antimetastatic; anti-angiogenic; antimalarial; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 19..1014
FT /*tag= a
FT /product= "NAF-1"
FT sig_peptide 19..87
FT /*tag= b
FT mat_peptide 88..1011
FT /*tag= c
XX
PN US5871969-A.
XX
PD 16-FEB-1999.
XX
PF 12-FEB-1997; 97US-0799173.
XX

PR 12-FEB-1996; 96US-0011519.
PR 12-FEB-1997; 97US-0799173.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Dillon PJ, Hastings G;
XX
DR WPI; 1999-166644/14.
DR P-PSDB; AAW92460.
XX
PT Nucleic acid encoding human neuronal attachment factor-1 - used to
PT e.g. treat spinal cord injuries, and inhibit tumor cell metastasis
PT and neovascularization
XX
PS Claim 2; Fig 1; 29pp; English.
XX
CC This sequence encodes a novel human neuronal attachment factor-1, NAF-1.
CC NAF-1 is an analogue of rat F-spondin and is used to treat spinal cord
CC injuries and damage to peripheral nerves (by promoting neural cell
CC adhesion and neurite extension), to inhibit tumor cell metastasis
CC (particularly in small cell and breast carcinoma) and endothelial cell
CC proliferation, adhesion and mobility, to reduce tumor neovascularisation,
CC as angiostatic agents for tumor cells, to promote wound healing and to
CC modulate hemostasis. NAF-1 may also be used to identify treatments and
CC diagnoses for human disease. Fragments of the nucleic acid that do not
CC encode NAF-1 peptides are useful as probes to isolate the NAF-1 gene, its
CC allelic variants, full-length cDNA or related sequences, in chromosomal
CC location by in situ hybridization or in Northern blotting, and as
CC diagnostic probes or primers. The protein has neurotrophic, anticancer,
CC antimetastatic, anti-angiogenic, antimalarial and modulates adhesion,
CC proliferation and mobility of cells.
XX
SQ Sequence 1105 BP; 178 A; 393 C; 362 G; 172 T; 0 other;

Query Match 100.0%; Score 1105; DB 20; Length 1105;
Best Local Similarity 100.0%; Pred. No. 3.4e-197;
Matches 1105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60

Qy      61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120

Qy     121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180

Qy     181 gccttccccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 gccttccccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggg 240

Qy     241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300

Qy     301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360

Qy     361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420

Qy     421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480

Qy     481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540

Qy     541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600

Qy     601 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660

Qy     661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720

Qy     721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780

Qy     781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840

Qy     841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900

Qy     901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960

Qy     961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020

Qy    1021 agccccgcagcccctggggcccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 agccccgcagcccctggggcccccccggagccatggggtgtcgggggctcctgtgcaggct 1080

Qy    1081 catgctgcaggcggccgagggcaca 1105
          ||||||||||||||||||||||||||
Db    1081 catgctgcaggcggccgagggcaca 1105
```

RESULT 2
AAT78360
ID AAT78360 standard; DNA; 1105 BP.
XX
AC AAT78360;
XX
DT 08-MAR-1998 (first entry)
XX
DE Human neuronal attachment factor-1 DNA.
XX
KW Neuronal attachment factor-1; NAF-1; human; cell-cell interaction;
KW cell adhesion; spinal cord injury; peripheral nerve damage; tumour;
KW metastasis; gene therapy; malaria; wound healing; ss.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	CDS	19..1014
FT		/*tag= a
FT	sig_peptide	19..87
FT		/*tag= b
FT	mat_peptide	88..1011
FT		/*tag= c

XX
PN WO9729189-A1.
XX
PD 14-AUG-1997.
XX
PF 12-FEB-1996; 96WO-US01857.
XX

PR 12-FEB-1996; 96WO-US01857.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Hastings GA;
XX
DR WPI; 1997-415344/38.
DR P-PSDB; AAW23663.
XX
PT New isolated human neuronal attachment factor-1 - promotes cell-cell
PT interaction and cell adhesion, used for treating e.g. spinal cord
PT injury, peripheral nerve damage or tumours
XX
PS Claim 5; Page 42-43; 55pp; English.
XX
CC This DNA sequence codes for novel human neuronal attachment
CC factor-1 (NAF-1) (see AAW23663), a protein that promotes cell-cell
CC interaction and cell adhesion. It was isolated from a cDNA library
CC derived from human epithelial sarcoma. NAF-1 DNA can be used in
CC the production of recombinant NAF-1 polypeptides (especially the
CC mature protein), in gene therapy methods for treatment of a patient
CC having need of NAF-1, and in the diagnosis of a disease or a
CC susceptibility to a disease related to an under-expression of NAF-1
CC resulting from gene mutation. NAF-1 can be used e.g. to treat
CC spinal cord injuries or damage to peripheral nerves by promoting
CC neural cell adhesion and neurite extension, to inhibit tumour cell
CC metastases, inhibit endothelial cell proliferation, adhesion and
CC motility, to decrease tumour neovascularisation, to be angiostatic
CC for tumours and to promote wound healing. It can also be used to
CC modulate haemostasis. Antagonists can be used to treat malaria.
XX
SQ Sequence 1105 BP; 178 A; 392 C; 363 G; 172 T; 0 other;

Query Match 99.6%; Score 1100.2; DB 18; Length 1105;
Best Local Similarity 99.7%; Pred. No. 2.7e-196;
Matches 1102; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60

Qy     61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120

Qy    121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180

Qy    181 gccttccccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggg 240
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    181 gccttccccaagcagtacccccctgttccgcccccctgcgcagtggtcttcgctgctgggg 240

Qy    241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300

Qy    301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360

Qy    361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420

Qy    421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480

Qy    481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540

Qy    541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600

Qy    601 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660

Qy    661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720

Qy    721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780

Qy    781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 gccccagtcctcgccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840

Qy    841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900

Qy    901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960

Qy    961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020

Qy   1021 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1080

Qy   1081 catgctgcaggcggccgagggcaca 1105
          ||||||||||||||||||||||||||
Db   1081 catgctgcaggcggccgagggcaca 1105
```

RESULT 3
AAZ95034
ID AAZ95034 standard; cDNA; 1840 BP.
XX
AC AAZ95034;
XX
DT 15-AUG-2000 (first entry)
XX
DE Cancer specific gene Pro108 cDNA useful as prostate cancer marker.
XX
KW Cancer specific gene; CSG; prostate cancer; diagnosis; monitoring;
KW staging; imaging; metastasis; therapy; human; Pro108; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 299..1294
FT /*tag= a
XX
PN WO200023108-A1.
XX
PD 27-APR-2000.
XX
PF 18-OCT-1999; 99WO-US23764.
XX
PR 19-OCT-1998; 98US-0104741.
XX
PA (DIAD-) DIADEXUS LLC.
XX
PI Ali SM, Sun Y, Salceda S, Recipon H, Cafferkey R;
XX

```
DR   WPI; 2000-339528/29.
DR   P-PSDB; AAY79561.
XX
PT   Diagnosing, detecting, staging, monitoring, imaging and treating
PT   cancers, especially useful for detecting prostate cancer comprises
PT   measuring changes in levels of cancer specific genes in cells, tissues
PT   and body fluids -
XX
PS   Claim 6; Page 28-29; 35pp; English.
XX
CC   The present sequence is that of a cancer specific gene (CSG)
CC   designated pro108. The CSG was identified using a suppression
CC   subtractive hybridization method. It is a diagnostic marker
CC   for prostate cancer. In 11 different healthy tissues examined,
CC   pro108 mRNA expression was highest in ovary and uterus. Expression
CC   in healthy prostate was low. However, overexpression of CSG pro108
CC   was observed in 13 of 13 prostate cancer tissues examined. CSGs
CC   comprising pro108, its encoded polypeptide (see AAY79561), or an
CC   antibody that specifically binds CSG, are used in claimed methods
CC   for the diagnosis, detection, staging, monitoring, imaging and
CC   treatment of prostate cancer. The new methods provide earlier
CC   diagnosis for the presence and metastasis of prostate cancer, and
CC   can be used to determine if a cancer has metastasized, or to
CC   monitor the progress or stage of the disease when it has not
CC   metastasized.
XX
SQ   Sequence 1840 BP; 305 A; 628 C; 582 G; 325 T; 0 other;

  Query Match             99.6%;  Score 1100.2;  DB 21;  Length 1840;
  Best Local Similarity   99.7%;  Pred. No. 2.7e-196;
  Matches 1102;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
Db  281 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 340

Qy   61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
Db  341 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 400

Qy  121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
Db  401 tgttccgccagagccccggccaaatacagcatcaccttcacgggcaagtggagccagacg 460

Qy  181 gccttccccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggg 240
Db  461 gccttccccaagcagtacccccctgttccgcccccctgcgcagtggtcttcgctgctgggg 520

Qy  241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
Db  521 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 580

Qy  301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
Db  581 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 640

Qy  361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
Db  641 gaggcgctgcagagcgtgcacgaggtgttttcggcgcccgccgtccccagcggcaccggg 700

Qy  421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
Db  701 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 760

Qy  481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
Db  761 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 820

Qy  541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
Db  821 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 880

Qy  601 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
Db  881 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 940

Qy  661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
Db  941 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 1000

Qy  721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
Db 1001 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 1060

Qy  781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
Db 1061 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 1120

Qy  841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
Db 1121 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 1180

Qy  901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
Db 1181 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 1240

Qy  961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
Db 1241 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1300

Qy 1021 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
Db 1301 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1360

Qy 1081 catgctgcaggcggccgagggcaca 1105
Db 1361 catgctgcaggcggccgagggcaca 1385
```

```
RESULT   4
AAH34981
ID   AAH34981 standard; cDNA; 1848 BP.
XX
AC   AAH34981;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Human colon cancer antigen encoding cDNA SEQ ID NO:2063.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US26524.
XX
PR   29-SEP-1999;   99US-0157137.
PR   03-NOV-1999;   99US-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Barash SC,  Birse CE,  Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   P-PSDB; AAG75576.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS   Claim 1; Page 3549; 9803pp; English.
XX
```

```
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer-associated Ps,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC   and AAB77789 represent sequences used in the exemplification of the
CC   present invention.
CC   N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.

     Sequence 1848 BP; 324 A; 601 C; 578 G; 340 T; 5 other;

  Query Match             99.5%;  Score 1100;  DB 22;  Length 1848;
  Best Local Similarity   99.5%;  Pred. No. 2.9e-196;
  Matches 1100;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 315

Qy      61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     316 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 375

Qy     121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     376 tgttccgccagagccccggccaaatacagcatcaccttcacgggcaagtggagccagacg 435

Qy     181 gccttccccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggg 240
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     436 gccttccccaagcagtacccccctgttccgcccccctgcgcagtggtcttcgctgctgggg 495

Qy     241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     496 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 555

Qy     301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     556 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 615

Qy     361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
           ||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db     616 gaggcgctgcagagcgtgcacgmggtgttttcggcgcccgccgtccccagcggcaccggg 675

Qy     421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     676 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 735

Qy     481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     736 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 795

Qy     541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     796 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 855

Qy     601 ttcaccttctcctccccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     856 ttcaccttctcctccccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 915

Qy     661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     916 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggcnctgcct 975

Qy     721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
           ||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db     976 cccatcgccagggtgacactgstgcggctgcgacagagccccagggccttcatccctccc 1035

Qy     781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1036 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 1095

Qy     841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1096 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 1155

Qy     901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1156 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 1215

Qy     961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1216 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1275

Qy    1021 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1276 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1335

Qy    1081 catgctgcaggcggccgagggcaca 1105
           ||||||||||||||||||||||||||
Db    1336 catgctgcaggcggccgagggcaca 1360


RESULT   5
AAV63241
ID   AAV63241 standard; cDNA; 1607 BP.
XX
AC   AAV63241;
XX
DT   21-JAN-1999  (first entry)
XX
DE   cDNA encoding an adhesion-modulating protein zsig25.
XX
KW   zsig25; adhesion-modulating protein; prostate cell; prostatic carcinoma;
KW   B-cell cancer; infertility; Wolf-Hirschhorn syndrome;
KW   chromosome 4 (p16.3); ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             40..1035
FT                   /*tag=  a
FT                   /product= zsig25
XX
PN   WO9845442-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US07117.
XX
PR   11-JUN-1997;   97US-0049288.
PR   10-APR-1997;   97US-0043421.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Sheppard PO;
XX
DR   WPI; 1998-557522/47.
DR   P-PSDB; AAW70589.
XX
PT   New zsig25 protein and related nucleic acid, fusion proteins,
PT   vectors, transformed cells - and antibodies, involved in modulation
PT   of adhesion, used for diagnosis and treatment of prostatic and
PT   B-cell tumours, stimulation of haematopoietic cells, treatment of
PT   immune deficiency etc.
```

```
XX
PS   Claim 40; Pages 108-110; 161pp; English.
XX
CC   The present sequence encodes a protein designated zsig25. The zsig25
CC   protein is an adhesion-modulating protein expressed at very high level
CC   in prostate cells. The protein is useful as a diagnostic marker for
CC   prostatic carcinoma and B-cell cancers, possibly also for infertility,
CC   and as a reagent for separating cancerous and non-cancerous cells.
CC   The products may also be used to diagnose or treat Wolf-Hirschhorn
CC   syndrome, associated with a deletion in the region of chromosome 4
CC   (p16.3) where the zsig25 gene is located.
XX
SQ   Sequence 1607 BP; 284 A; 540 C; 494 G; 289 T; 0 other;


  Query Match             99.4%;  Score 1098.6;  DB 19;  Length 1607;
  Best Local Similarity   99.6%;  Pred. No. 5.3e-196;
  Matches 1101;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     22 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 81

Qy     61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     82 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 141

Qy    121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    142 tgttccgccagagccccggccaaatacagcatcaccttcacgggcaagtggagccagacg 201

Qy    181 gccttccccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggg 240
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    202 gccttccccaagcagtacccccctgttccgcccccctgcgcagtggtcttcgctgctgggg 261

Qy    241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    262 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 321

Qy    301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    322 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 381

Qy    361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    382 gaggcgctgcagagcgtgcacgaggtgttttcggcgcccgccgtccccagcggcaccggg 441

Qy    421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    442 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 501

Qy    481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    502 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 561

Qy    541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    562 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 621

Qy    601 ttcaccttctcctccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    622 ttcaccttctcctccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 681

Qy    661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    682 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 741

Qy    721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    742 cccatcgccagggtgacactgctgcggctgcgacagagccccagggccttcatccctccc 801

Qy    781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    802 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 861

Qy    841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    862 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 921

Qy    901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    922 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 981

Qy    961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    982 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1041

Qy   1021 agccccgcagcccctggggcccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1042 agccccgcagcccctggggcccccccggagccatggggtgtcgggggctcctgtgcaggct 1101

Qy   1081 catgctgcaggcggccgagggcaca 1105
          ||||||||||||||||||||||||
Db   1102 catgctgcaggcggccgagggcaca 1126


RESULT   6
AAF90567
ID   AAF90567 standard; DNA; 1770 BP.
XX
AC   AAF90567;
XX
DT   22-AUG-2001  (first entry)
XX
DE   Human extracellular matrix protein RG1 DNA.
XX
KW   RG1; human; extracellular matrix protein; prostate cancer;
KW   metastasis; tumour; benign prostatic hyperplasia; gene therapy;
KW   diagnosis; antitumour; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             296..1294
FT                   /*tag= a
XX
PN   WO200144291-A2.
XX
PD   21-JUN-2001.
XX
PF   15-DEC-2000; 2000WO-US33901.
XX
PR   16-DEC-1999;   99US-0172370.
PR   07-DEC-2000; 2000US-9966561.
XX
PA   (SCHD ) SCHERING AG.
XX
PI   Harkins R,  Parkes D,  Parry G,  Schneider DW,  Steinbrecher R;
XX
DR   WPI; 2001-398128/42.
DR   P-PSDB; AAB90567.
XX
PT   Novel human extracellular matrix polypeptide, RG1, useful in research,
PT   diagnosis and treatment of metastasis such as prostate cancer
XX
PS   Disclosure; Fig 4; 69pp; English.
XX
CC   The present sequence is that of a polynucleotide, designated rg1,
CC   encoding novel human extracellular matrix polypeptide RG1 (see
CC   AAB82472). The full-length coding sequence was deduced from a
CC   clone identified from a database screening that was found in the
CC   category of cell adhesion molecules and described as a homologue
CC   of f-spondin. RG1 is expressed in prostate tissue and may be
CC   over-expressed in prostate tumours. It shows homology to a
CC   superfamily of extracellular matrix proteins encoded by the
CC   mindin/f-spondin genes. The invention provides human RG1
```

```
CC   polypeptides, polynucleotides encoding them (especially encoding
CC   amino acids 1-331, 21-331 or 27-331 of RG1), methods for
CC   producing the polypeptides, expression vectors and genetically
CC   engineered host cells for expression of the polypeptides.  It
CC   also provides methods for utilizing the polynucleotides and
CC   polypeptides in research, diagnosis and therapeutic applications.
CC   Thus, rg1 polynucleotides can be used in the production of
CC   recombinant proteins, in methods for assessing RG1 expression
CC   in cells, and in the development of ribozyme and antisense
CC   oligonucleotides useful for administration e.g. to a human
CC   patient for treatment of a disease such as prostate cancer or
CC   benign prostatic hyperplasia, which is alleviated by decreasing
CC   the level of RG1 activity.
XX
SQ   Sequence 1770 BP; 285 A; 614 C; 570 G; 301 T; 0 other;


  Query Match             99.4%;  Score 1098.6;  DB 22;  Length 1770;
  Best Local Similarity   99.6%;  Pred. No. 5.3e-196;
  Matches 1101;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     278 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 337

Qy      61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     338 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 397

Qy     121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
           ||||||||| |||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     398 tgttccgccggagccccggccaaatacagcatcaccttcacgggcaagtggagccagacg 457

Qy     181 gccttccccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggg 240
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     458 gccttccccaagcagtacccccctgttccgcccccctgcgcagtggtcttcgctgctgggg 517

Qy     241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     518 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 577

Qy     301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     578 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 637

Qy     361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     638 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 697

Qy     421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     698 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 757

Qy     481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     758 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 817

Qy     541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     818 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 877

Qy     601 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     878 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 937

Qy     661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     938 tcctcctctcccagccacccggccaactccttctactacccacggctgaaggccctgcct 997

Qy     721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     998 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 1057

Qy     781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1058 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 1117

Qy     841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1118 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 1177

Qy     901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1178 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 1237

Qy     961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1238 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1297

Qy    1021 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1298 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1357

Qy    1081 catgctgcaggcggccgagggcaca 1105
           ||||||||||||||||||||||||||
Db    1358 catgctgcaggcggccgagggcaca 1382


RESULT   7
AAF90566
ID   AAF90566 standard; DNA; 1785 BP.
XX
AC   AAF90566;
XX
DT   22-AUG-2001  (first entry)
XX
DE   Human extracellular matrix protein RG1 DNA.
XX
KW   RG1; human; extracellular matrix protein; prostate cancer;
KW   metastasis; tumour; benign prostatic hyperplasia; gene therapy;
KW   diagnosis; antitumour; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             296..1294
FT                   /*tag=  a
XX
PN   WO200144291-A2.
XX
PD   21-JUN-2001.
XX
PF   15-DEC-2000; 2000WO-US33901.
XX
PR   16-DEC-1999;   99US-0172370.
PR   07-DEC-2000; 2000US-9966561.
XX
PA   (SCHD ) SCHERING AG.
XX
PI   Harkins R,  Parkes D,  Parry G,  Schneider DW,  Steinbrecher R;
XX
DR   WPI; 2001-398128/42.
DR   P-PSDB; AAB82472.
XX
PT   Novel human extracellular matrix polypeptide, RG1, useful in research,
PT   diagnosis and treatment of metastasis such as prostate cancer  -
XX
PS   Claim 8; Fig 1; 69pp; English.
XX
CC   The present sequence is that of a polynucleotide, designated rg1,
CC   encoding novel human extracellular matrix polypeptide RG1 (see
CC   AAB82472).  The full-length coding sequence was deduced from a
CC   clone identified from a database screening that was found in the
CC   category of cell adhesion molecules and described as a homologue
CC   of f-spondin.  RG1 is expressed in prostate tissue and may be
CC   over-expressed in prostate tumours.  It shows homology to a
```

```
CC   superfamily of extracellular matrix proteins encoded by the
CC   mindin/f-spondin genes.  The invention provides human RG1
CC   polypeptides, polynucleotides encoding them (especially encoding
CC   amino acids 1-331, 21-331 or 27-331 of RG1), methods for
CC   producing the polypeptides, expression vectors and genetically
CC   engineered host cells for expression of the polypeptides.  It
CC   also provides methods for utilizing the polynucleotides and
CC   polypeptides in research, diagnosis and therapeutic applications.
CC   Thus, rg1 polynucleotides can be used in the production of
CC   recombinant proteins, in methods for assessing RG1 expression
CC   in cells, and in the development of ribozyme and antisense
CC   oligonucleotides useful for administration e.g. to a human
CC   patient for treatment of a disease such as prostate cancer or
CC   benign prostatic hyperplasia, which is alleviated by decreasing
CC   the level of RG1 activity.
XX
SQ   Sequence 1785 BP; 300 A; 614 C; 570 G; 301 T; 0 other;


  Query Match             99.4%;  Score 1098.6;  DB 22;  Length 1785;
  Best Local Similarity   99.6%;  Pred. No. 5.3e-196;
  Matches 1101;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     278 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 337

Qy      61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     338 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 397

Qy     121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
           |||||||||| |||||| ||||||||||||||||||||||||||||||||||||||||||
Db     398 tgttccgccggagccccggccaaatacagcatcaccttcacgggcaagtggagccagacg 457

Qy     181 gccttccccaagcagtacccccctgttccgccccccctgcccagtggtcttcgctgctgggg 240
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     458 gccttccccaagcagtacccccctgttccgccccccctgcgcagtggtcttcgctgctgggg 517

Qy     241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     518 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 577

Qy     301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     578 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 637

Qy     361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     638 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 697

Qy     421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     698 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 757

Qy     481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     758 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 817

Qy     541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     818 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 877

Qy     601 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     878 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 937

Qy     661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     938 tcctcctctcccagccacccggccaactccttctactacccacggctgaaggccctgcct 997

Qy     721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     998 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 1057

Qy     781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1058 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 1117

Qy     841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1118 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 1177

Qy     901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1178 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 1237

Qy     961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1238 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1297

Qy    1021 agccccgcagcccctggggcccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1298 agccccgcagcccctggggcccccccggagccatggggtgtcgggggctcctgtgcaggct 1357

Qy    1081 catgctgcaggcggccgagggcaca 1105
           |||||||||||||||||||||||||
Db    1358 catgctgcaggcggccgagggcaca 1382


RESULT   8
AAZ34089
ID   AAZ34089 standard; cDNA; 1846 BP.
XX
AC   AAZ34089;
XX
DT   07-DEC-1999  (first entry)
XX
DE   Human PRO866 nucleotide sequence.
XX
KW   Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW   probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW   secreted protein; transmembrane protein; ss.
XX
OS   Homo sapiens.
XX
PN   WO9946281-A2.
XX
PD   16-SEP-1999.
XX
PF   08-MAR-1999;   99WO-US05028.
XX
PR   10-MAR-1998;   98US-0077450.
PR   11-MAR-1998;   98US-0077632.
PR   11-MAR-1998;   98US-0077641.
PR   11-MAR-1998;   98US-0077649.
PR   12-MAR-1998;   98US-0077791.
PR   13-MAR-1998;   98US-0078004.
PR   17-MAR-1998;   98US-0040220.
PR   20-MAR-1998;   98US-0078886.
PR   20-MAR-1998;   98US-0078910.
PR   20-MAR-1998;   98US-0078936.
PR   20-MAR-1998;   98US-0078939.
PR   25-MAR-1998;   98US-0079294.
PR   26-MAR-1998;   98US-0079656.
PR   27-MAR-1998;   98US-0079663.
PR   27-MAR-1998;   98US-0079664.
PR   27-MAR-1998;   98US-0079689.
PR   27-MAR-1998;   98US-0079728.
PR   27-MAR-1998;   98US-0079786.
PR   30-MAR-1998;   98US-0079920.
PR   30-MAR-1998;   98US-0079923.
PR   31-MAR-1998;   98US-0080105.
PR   31-MAR-1998;   98US-0080107.
PR   31-MAR-1998;   98US-0080165.
PR   31-MAR-1998;   98US-0080194.
```

PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 08-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 09-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081838.
PR 15-APR-1998; 98US-0081952.
PR 15-APR-1998; 98US-0081955.
PR 21-APR-1998; 98US-0082568.
PR 21-APR-1998; 98US-0082569.
PR 22-APR-1998; 98US-0082700.
PR 22-APR-1998; 98US-0082704.
PR 22-APR-1998; 98US-0082804.
PR 23-APR-1998; 98US-0082767.
PR 23-APR-1998; 98US-0082796.
PR 27-APR-1998; 98US-0083336.
PR 28-APR-1998; 98US-0083322.
PR 29-APR-1998; 98US-0083392.
PR 29-APR-1998; 98US-0083495.
PR 29-APR-1998; 98US-0083496.
PR 29-APR-1998; 98US-0083499.
PR 29-APR-1998; 98US-0083500.
PR 29-APR-1998; 98US-0083545.
PR 29-APR-1998; 98US-0083554.
PR 29-APR-1998; 98US-0083558.
PR 29-APR-1998; 98US-0083559.
PR 30-APR-1998; 98US-0083742.
PR 05-MAY-1998; 98US-0084366.
PR 06-MAY-1998; 98US-0084414.
PR 06-MAY-1998; 98US-0084441.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.
PR 07-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085338.
PR 13-MAY-1998; 98US-0085339.
PR 15-MAY-1998; 98US-0085573.
PR 15-MAY-1998; 98US-0085579.
PR 15-MAY-1998; 98US-0085580.
PR 15-MAY-1998; 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.
PR 15-MAY-1998; 98US-0085700.
PR 15-MAY-1998; 98US-0085704.
PR 18-MAY-1998; 98US-0086023.
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086486.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.
XX
PA (GETH) GENENTECH INC.
XX
PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
XX
DR WPI; 1999-551358/46.
DR P-PSDB; AAY41721.
XX
PT New secreted and transmembrane polypeptides and their polynucleotides,
PT useful for treating blood coagulation disorders, cancers and cellular
PT adhesion disorders -
XX
PS Claim 2; Fig 86; 530pp; English.
XX
CC The present invention describes secreted and transmembrane polypeptides
CC and their polynucleotides. The nucleotide sequences are useful as
CC sources of probes, primers, for chromosome mapping, and for generation
CC of antisense sequences. They can also be used to create transgenic
CC animals. The proteins can be used to treat a variety of diseases and
CC disorders, depending on their function. Diseases that may be treated
CC include blood coagulation disorders, cancers and cellular adhesion
CC disorders. They may also be used to raise antibodies. AAZ33891 to
CC AAZ34338, and AAY41685 to AAY41774 represent polynucleotide and
CC polypeptide sequence given in the exemplification of the present
CC invention.
XX
SQ Sequence 1846 BP; 331 A; 622 C; 571 G; 322 T; 0 other;

Query Match 99.4%; Score 1098.6; DB 20; Length 1846;
Best Local Similarity 99.6%; Pred. No. 5.3e-196;
Matches 1101; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy      1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    257 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 316

Qy     61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    317 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 376

Qy    121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    377 tgttccgccagagccccggccaaatacagcatcaccttcacgggcaagtggagccagacg 436

Qy    181 gccttccccaagcagtacccccctgttccgccccccctgcccagtggtcttcgctgctgggg 240
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    437 gccttccccaagcagtacccccctgttccgccccccctgcgcagtggtcttcgctgctgggg 496

Qy    241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    497 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 556

Qy    301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    557 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 616

Qy    361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    617 gaggcgctgcagagcgtgcacgaggtgttttcggcgcccgccgtccccagcggcaccggg 676

Qy    421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    677 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 736

Qy    481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    737 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 796

Qy    541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    797 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 856

Qy    601 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    857 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 916

Qy    661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    917 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 976

Qy    721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
```

```
Db    977 cccatcgccagggtgacactgctgcggctgcgacagagccccagggccttcatccctccc 1036

Qy    781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1037 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 1096

Qy    841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1097 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 1156

Qy    901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1157 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 1216

Qy    961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1217 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1276

Qy   1021 agccccgcagcccctggggcccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1277 agccccgcagcccctggggcccccccggagccatggggtgtcgggggctcctgtgcaggct 1336

Qy   1081 catgctgcaggcggccgagggcaca 1105
          |||||||||||||||||||||||||
Db   1337 catgctgcaggcggccgagggcaca 1361
```

```
RESULT   9
AAC78507
ID   AAC78507 standard; cDNA; 1847 BP.
XX
AC   AAC78507;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human PRO866 (UNQ435) nucleotide sequence SEQ ID NO:235.
XX
KW   Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
KW   expressed sequence tag; detection; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200053756-A2.
XX
PD   14-SEP-2000.
XX
PF   18-FEB-2000; 2000WO-US04341.

     08-MAR-1999;   99WO-US05028.
     12-MAR-1999;   99US-0123957.
PR   29-MAR-1999;   99US-0126773.
PR   21-APR-1999;   99US-0130232.
PR   28-APR-1999;   99US-0131445.
PR   14-MAY-1999;   99US-0134287.
PR   23-JUN-1999;   99US-0141037.
PR   26-JUL-1999;   99US-0145698.
PR   29-OCT-1999;   99US-0162506.
PR   30-NOV-1999;   99WO-US28313.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28565.
PR   16-DEC-1999;   99WO-US30095.
PR   30-DEC-1999;   99WO-US31243.
PR   30-DEC-1999;   99WO-US31274.
PR   05-JAN-2000;  2000WO-US00219.
PR   06-JAN-2000;  2000WO-US00277.
PR   06-JAN-2000;  2000WO-US00376.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Botstein D,  Desnoyers L,  Eaton DL;
PI   Ferrara N,  Filvaroff E,  Fong S,  Gao W,  Gerber H,  Gerritsen ME;
PI   Goddard A,  Godowski PJ,  Grimaldi CJ,  Gurney AL,  Hillan KJ;
```

```
PI   Kljavin IJ,  Kuo SS,  Napier MA,  Pan J,  Paoni NF,  Roy MA;
PI   Shelton DL,  Stewart TA,  Tumas D,  Williams PM,  Wood WI;
XX
DR   WPI; 2000-611443/58.
DR   P-PSDB; AAB44277.
XX
PT   Novel PRO polypeptides and polynucleotides used in detection methods,
PT   to target bioactive molecules to specific cells, and to modulate
PT   cellular activities -
XX
PS   Claim 2; Fig 86; 636pp; English.
XX
CC   AAC78458 to AAC78599 represent polynucleotide and EST (expressed
CC   sequence tag) sequences which encode secreted or transmembrane PRO
CC   polypeptides. The PRO polynucleotides and polypeptides have cytostatic
CC   activity. The polynucleotides and polypeptides can be used for detecting
CC   the presence of PRO polypeptides in samples, for linking bioactive
CC   molecules to cells and for modulating biological activities of cells,
CC   using the polypeptides for specific targeting. The polypeptide targeting
CC   can be used to kill the target cells, e.g. for the treatment of cancers.
CC   The polypeptide pairs provide specific targeting of bioactive molecules
CC   to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
CC   the isolation of the PRO polynucleotide sequences.
XX
SQ   Sequence 1847 BP; 332 A; 622 C; 571 G; 322 T; 0 other;
```

```
  Query Match             99.4%;  Score 1098.6;  DB 21;  Length 1847;
  Best Local Similarity   99.6%;  Pred. No. 5.3e-196;
  Matches 1101;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    257 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 316

Qy     61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    317 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 376

Qy    121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    377 tgttccgccagagccccggccaaatacagcatcaccttcacgggcaagtggagccagacg 436

Qy    181 gccttccccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggg 240
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    437 gccttccccaagcagtacccccctgttccgcccccctgcgcagtggtcttcgctgctgggg 496

Qy    241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    497 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 556

Qy    301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    557 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 616

Qy    361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    617 gaggcgctgcagagcgtgcacgaggtgttttcggcgcccgccgtccccagcggcaccggg 676

Qy    421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    677 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 736

Qy    481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    737 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 796

Qy    541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    797 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 856

Qy    601 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    857 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 916

Qy    661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    917 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 976

Qy    721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    977 cccatcgccagggtgacactgctgcggctgcgacagagccccagggccttcatccctccc 1036

Qy    781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1037 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 1096

Qy    841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1097 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 1156

Qy    901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1157 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 1216

Qy    961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1217 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1276

Qy   1021 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1277 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1336

Qy   1081 catgctgcaggcggccgagggcaca 1105
          |||||||||||||||||||||||||
Db   1337 catgctgcaggcggccgagggcaca 1361
```

RESULT 10
AAC58630
ID AAC58630 standard; cDNA; 1847 BP.
XX
AC AAC58630;
XX
DT 29-JAN-2001 (first entry)
XX
DE Human PRO866 protein UNQ435 encoding cDNA SEQ ID NO:257.
XX
KW Human; immune related disease; diagnosis; antiinflammatory; cardiant;
KW dermatological; antiarthritic; antirheumatic; immunosuppressive;
KW haemostatic; antithyroid; antidiabetic; nootropic; neuroprotective;
KW antianaemic; hepatotropic; virucide; antipsoriatic; antiallergic;
KW antiasthmatic; systemic lupus erythematosus; rheumatoid arthritis;
KW osteoarthritis; spondyloarthropathy; systemic sclerosis; sarcoidosis;
KW idiopathic inflammatory myopathy; Sjogren's syndrome; thyroiditis;
KW systemic vasculitis; autoimmune haemolytic anaemia; diabetes mellitus;
KW autoimmune thrombocytopaenia; immune-mediated renal disease;
KW demyelinating disease; hepatobiliary disease; Whipple's disease;
KW inflammatory bowel disease; gluten-sensitive enteropathy;
KW autoimmune disease; immune-mediated skin disease; allergic disease;
KW immunological disease; transplantation associated disease;
KW graft rejection; graft-versus-host-disease; ss.
XX
OS Homo sapiens.
XX
PN WO200053758-A2.
XX
PD 14-SEP-2000.
XX
PF 02-MAR-2000; 2000WO-US05841.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 10-MAR-1999; 99US-0123618.
PR 12-MAR-1999; 99US-0123957.
PR 23-MAR-1999; 99US-0125775.
PR 12-APR-1999; 99US-0128849.
PR 20-APR-1999; 99WO-US08615.
PR 28-APR-1999; 99US-0131445.
PR 04-MAY-1999; 99US-0132371.
PR 14-MAY-1999; 99US-0134287.
PR 02-JUN-1999; 99WO-US12252.
PR 23-JUN-1999; 99US-0141037.
PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 28-JUL-1999; 99US-0146222.
PR 01-SEP-1999; 99WO-US20111.
PR 08-SEP-1999; 99WO-US20594.
PR 13-SEP-1999; 99WO-US20944.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 05-OCT-1999; 99WO-US23089.
PR 29-OCT-1999; 99US-0162506.
PR 29-NOV-1999; 99WO-US28214.
PR 30-NOV-1999; 99WO-US28313.
PR 30-NOV-1999; 99WO-US28409.
PR 01-DEC-1999; 99WO-US28301.
PR 01-DEC-1999; 99WO-US28634.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28564.
PR 02-DEC-1999; 99WO-US28565.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30999.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
XX
PA (GETH) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W;
PI Kabakoff RC, Lu Y, Pan J, Pennica D, Shelton DL, Smith V;
PI Stewart TA, Tumas D, Watanabe CK, Wood WI, Yan M;
XX
DR WPI; 2000-572271/53.
DR P-PSDB; AAB33465.
XX
PT Sixty four PRO polypeptides, useful in the diagnosis and treatment of
PT immune related disorders, e.g. systemic lupus erythematosis, rheumatoid
PT arthritis, osteoarthritis, thyroiditis and diabetes mellitus -
XX
PS Claim 23; Fig 103; 309pp; English.
XX
CC The present invention describes sixty four human PRO proteins which can
CC be used in the treatment of immune related diseases. The human PRO
CC proteins, anti-PRO antibodies, agonists and antagonists are useful for
CC treating and diagnosing immune related disorders. The disorders are
CC selected from systemic lupus erythematosus, rheumatoid arthritis,
CC osteoarthritis, juvenile chronic arthritis, spondyloarthropathies,
CC systemic sclerosis, idiopathic inflammatory myopathies, Sjogren's
CC syndrome, systemic vasculitis, sarcoidosis, autoimmune haemolytic
CC anaemia, autoimmune thrombocytopaenia, thyroiditis, diabetes mellitus,
CC immune-mediated renal disease, demyelinating diseases of the central
CC and peripheral nervous systems, hepatobiliary diseases, inflammatory
CC bowel disease, gluten-sensitive enteropathy and Whipple's disease,
CC autoimmune or immune-mediated skin diseases, allergic diseases,
CC immunological diseases of the lung, and transplantation associated
CC diseases including graft rejection and graft-versus-host-disease.
CC AAC58397 to AAC58578 represent PCR primers and hybridisation probes used
CC in the isolation of human PRO sequences. AAC58579 to AAC58642 and
CC AAB33414 to AAB33477 represent human PRO polynucleotide and protein
CC sequences given in the exemplification of the present invention.
XX
SQ Sequence 1847 BP; 332 A; 622 C; 571 G; 322 T; 0 other;

```
  Query Match            99.4%;  Score 1098.6;  DB 21;  Length 1847;
  Best Local Similarity  99.6%;  Pred. No. 5.3e-196;
  Matches 1101;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     257 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 316

Qy      61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 376

Qy     121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     377 tgttccgccagagccccggccaaatacagcatcaccttcacgggcaagtggagccagacg 436

Qy     181 gccttccccaagcagtacccccctgttccgccccccctgcccagtggtcttcgctgctgggg 240
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     437 gccttccccaagcagtacccccctgttccgccccccctgcgcagtggtcttcgctgctgggg 496

Qy     241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     497 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 556

Qy     301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     557 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 616

Qy     361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     617 gaggcgctgcagagcgtgcacgaggtgttttcggcgcccgccgtccccagcggcaccggg 676

Qy     421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     677 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 736

Qy     481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     737 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 796

Qy     541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     797 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 856

Qy     601 ttcaccttctcctccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     857 ttcaccttctcctccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 916

Qy     661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     917 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 976

Qy     721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     977 cccatcgccagggtgacactgctgcggctgcgacagagccccagggccttcatccctccc 1036

Qy     781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1037 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 1096

Qy     841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1097 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 1156

Qy     901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1157 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 1216

Qy     961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1217 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1276

Qy    1021 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1277 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1336

Qy    1081 catgctgcaggcggccgagggcaca 1105
          |||||||||||||||||||||||||
Db    1337 catgctgcaggcggccgagggcaca 1361


RESULT  11
AAA49728
ID   AAA49728 standard; cDNA; 1847 BP.
XX
AC   AAA49728;
XX
DT   25-SEP-2000  (first entry)
XX
DE   Human PRO866 cDNA clone DNA53971-1359.
XX
KW   PRO866; human; antitumour; tumour; therapy; cytostatic;
KW   breast cancer; ovarian cancer; renal cancer; colorectal cancer;
KW   uterine cancer; prostate cancer; lung cancer; bladder cancer;
KW   central nervous system cancer; melanoma; leukaemia; neoplasm; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             275..1270
FT                   /*tag=  a
FT   sig_peptide     275..352
FT                   /*tag=  b
FT   mat_peptide     353..1267
FT                   /*tag=  c
XX
PN   WO200037638-A2.
XX
PD   29-JUN-2000.
XX
PF   02-DEC-1999;   99WO-US28565.
XX
PR   22-DEC-1998;   98US-0113296.
PR   08-MAR-1999;   99WO-US05028.
PR   21-APR-1999;   99US-0130232.
PR   28-APR-1999;   99US-0131445.
PR   14-MAY-1999;   99US-0134287.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Goddard A,  Godowski PJ,  Gurney AL,  Marsters SA;
PI   Napier MA,  Pitti RM,  Wood WI;
XX
DR   WPI; 2000-442668/38.
DR   P-PSDB; AAY95349.
XX
PT   Novel composition to inhibit neoplastic cell growth or for treating
PT   tumor in mammal comprises polypeptides PRO179, PRO207, PRO320, PRO219,
PT   PRO221, PRO224, PRO328, PRO301, PRO526, PRO362, PRO356, PRO509 or
PT   PRO866  -
XX
PS   Claim 20; Fig 25; 172pp; English.
XX
CC   The present sequence is that of cDNA clone DNA53971-1359
CC   (ATCC 209750) encoding human PRO866 (see AAY95349), a novel antitumour
CC   protein showing similarity to mindin/spondin proteins.  The cDNA
CC   was isolated from a human foetal kidney cDNA library by hybridisation.
CC   A claimed method for inhibiting the growth of a tumour cell comprises
CC   exposing the tumor cell to PRO179, PRO207, PRO320, PRO219, PRO221,
CC   PRO224, PRO328, PRO301, PRO526, PRO362, PRO356, PRO509 or PRO866 (see
CC   AAY95337-49), their agonists or chimeric polypeptides incorporating
CC   them.  The tumour is especially a cancer selected from breast,
```

```
CC   ovarian, renal, colorectal, uterine, prostate, lung, bladder and
CC   central nervous system cancer, melanoma and leukaemia.  Nucleic
CC   acids encoding PRO179 etc. are used in the recombinant production
CC   of the antitumour polypeptides.
XX
SQ   Sequence 1847 BP; 332 A; 622 C; 571 G; 322 T; 0 other;


  Query Match             99.4%;  Score 1098.6;  DB 21;  Length 1847;
  Best Local Similarity   99.6%;  Pred. No. 5.3e-196;
  Matches 1101;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     257 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 316

Qy      61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 376

Qy     121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     377 tgttccgccagagccccggccaaatacagcatcaccttcacgggcaagtggagccagacg 436

Qy     181 gccttccccaagcagtacccccctgttccgccccccctgcccagtggtcttcgctgctgggg 240
           ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     437 gccttccccaagcagtacccccctgttccgccccccctgcgcagtggtcttcgctgctgggg 496

Qy     241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     497 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 556

Qy     301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     557 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 616

Qy     361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     617 gaggcgctgcagagcgtgcacgaggtgttttcggcgcccgccgtccccagcggcaccggg 676

Qy     421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     677 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 736

Qy     481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     737 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 796

Qy     541 cgttggcgggaacaggcggcgctggacctgtaccctacgacgccgggacggacagcggc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     797 cgttggcgggaacaggcggcgctggacctgtaccctacgacgccgggacggacagcggc 856

Qy     601 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     857 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 916

Qy     661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     917 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 976

Qy     721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     977 cccatcgccagggtgacactgctgcggctgcgacagagccccagggccttcatccctccc 1036

Qy     781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1037 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 1096

Qy     841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1097 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 1156

Qy     901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1157 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 1216

Qy     961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1217 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1276

Qy    1021 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1277 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1336

Qy    1081 catgctgcaggcggccgagggcaca 1105
           ||||||||||||||||||||||||||
Db    1337 catgctgcaggcggccgagggcaca 1361


RESULT  12
AAI59814/c
ID   AAI59814 standard; cDNA; 1926 BP.
XX
AC   AAI59814;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 3803.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM40658.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 3803; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
```

```
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 1926 BP; 337 A; 604 C; 645 G; 334 T; 6 other;

  Query Match             99.4%;  Score 1098.6;  DB 22;  Length 1926;
  Best Local Similarity   99.6%;  Pred. No. 5.3e-196;
  Matches 1101;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1563 CGCTGCTCCTGCCGGGTGATGGAAAACCCCAGCCCGGCCGCCGCCCTGGGCAAGGCCCTC 1504

Qy      61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    1503 TGCGCTCTCCTCCTGGCCACTCTCGGCGCCGCCGGCCAGCCTCTTGGGGGAGAATCCATC 1444

Qy     121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1443 TGTTCCGCCAGAGCCCCGGCCAAATACAGCATCACCTTCACGGGCAAGTGGAGCCAGACG 1384

Qy     181 gccttccccaagcagtacccccctgttccgccccccctgcccagtggtcttcgctgctgggg 240
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    1383 GCCTTCCCCAAGCAGTACCCCCTGTTCCGCCCCCCTGCGCAGTGGTCTTCGCTGCTGGGG 1324

Qy     241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1323 GCCGCGCATAGCTCCGACTACAGCATGTGGAGGAAGAACCAGTACGTCAGTAACGGGCTG 1264

Qy     301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1263 CGCGACTTTGCGGAGCGCGGCGAGGCCTGGGCGCTGATGAAGGAGATCGAGGCGGCGGGG 1204

Qy     361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1203 GAGGCGCTGCAGAGCGTGCACGCGGTGTTTTCGGCGCCCGCCGTCCCCAGCGGCACCGGG 1144

Qy     421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1143 CAGACGTCGGCGGAGCTGGAGGTGCAGCGCAGGCACTCGCTGGTCTCGTTTGTGGTGCGC 1084

Qy     481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1083 ATCGTGCCCAGCCCCGACTGGTTCGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGGAC 1024

Qy     541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1023 CGTTGGCGGGAACAGGCGGCGCTGGACCTGTACCCCTACGACGCCGGGACGGACAGCGGC 964

Qy     601 ttcaccttctcctccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     963 TTCACCTTCTCCTCCCCCAACTTCGCCACCATCCCGCAGGACACGGTGACCGAGATAACG 904

Qy     661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     903 TCCTCCTCTCCCAGCCACCCGGCCAACTCCTTCTACTACCCGCGGCTGAAGGCCCTGCCT 844

Qy     721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     843 CCCATCGCCAGGGTGACACTGCTGCGGCTGCGACAGAGCCCCAGGGCCTTCATCCCTCCC 784

Qy     781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     783 GCCCCAGTCCTGCCCAGCAGGGACAATGAGATTGTAGACAGCGCCTCAGTTCCAGAAACG 724

Qy     841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     723 CCGCTGGACTGCGAGGTCTCCCTGTGGTCGTCCTGGGGACTGTGCGGAGGCCACTGTGGG 664

Qy     901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     663 AGGCTCGGGACCAAGAGCAGGACTCGCTACGTCCGGGTCCAGCCCGCCAACAACGGGAGC 604

Qy     961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     603 CCCTGCCCCGAGCTCGAAGAAGAGGCTGAGTGCGTCCCTGATAACTGCGTCTAAGACCAG 544

Qy    1021 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     543 AGCCCCGCAGCCCCTGGGGCCCCCCGGAGCCATGGGGTGTCGGGGGCTCCTGTGCAGGCT 484

Qy    1081 catgctgcaggcggccgagggcaca 1105
           |||||||||||||||||||||||||
Db     483 CATGCTGCAGGCGGCCGAGGGCACA 459


RESULT  13
AAI58028
ID   AAI58028 standard; cDNA; 1991 BP.
XX
AC   AAI58028;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 231.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM38872.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 231; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
```

```
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 1991 BP; 339 A; 669 C; 620 G; 350 T; 13 other;

  Query Match             99.4%;  Score 1098.6;  DB 22;  Length 1991;
  Best Local Similarity   99.6%;  Pred. No. 5.3e-196;
  Matches 1101;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     371 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 430

Qy      61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     431 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagaatccatc 490

Qy     121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     491 tgttccgccagagccccggccaaatacagcatcaccttcacgggcaagtggagccagacg 550

Qy     181 gccttccccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggg 240
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     551 gccttccccaagcagtacccccctgttccgcccccctgcgcagtggtcttcgctgctgggg 610

Qy     241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    -611 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 670

Qy     301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     671 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 730

Qy     361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     731 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 790

Qy     421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     791 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 850

Qy     481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     851 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 910

Qy     541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     911 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 970

Qy     601 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     971 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 1030

Qy     661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1031 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 1090

Qy     721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    1091 cccatcgccagggtgacactgctgcggctgcgacagagccccagggccttcatccctccc 1150

Qy     781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1151 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 1210

Qy     841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1211 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 1270

Qy     901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1271 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 1330

Qy     961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1331 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1390

Qy    1021 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1391 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1450

Qy    1081 catgctgcaggcggccgagggcaca 1105
           |||||||||||||||||||||||||
Db    1451 catgctgcaggcggccgagggcaca 1475


RESULT  14
AAC59794
ID   AAC59794 standard; DNA; 1718 BP.
XX
AC   AAC59794;
XX
DT   26-JAN-2001  (first entry)
XX
DE   Human secreted protein encoding DNA clone vp22 1.
XX
KW   Secreted protein; human; autoimmune disorder; multiple sclerosis; ulcer;
KW   systemic lupus erythematosus; rheumatoid arthritis; anaemia; stroke;
KW   haematopoiesis regulation; tissue regrowth; wound healing; haemophilia;
KW   Alzheimer's disease; Parkinson's disease; Shy-drager syndrome; cancer;
KW   contraceptive; infection; growth inhibition; hyperproliferative disorder;
KW   psoriasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200055375-A1.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US07285.
XX
PR   17-MAR-1999;   99US-0124808.
PR   17-MAR-1999;   99US-0124916.
PR   17-AUG-1999;   99US-0149639.
PR   01-OCT-1999;   99US-0157247.
PR   29-NOV-1999;   99US-0167824.
PR   15-FEB-2000;  2000US-0182711.
XX
PA   (ALPH-) ALPHAGENE INC.
XX
PI   Valenzuela D,  Yuan O,  Hoffman H,  Hall J,  Rapiejko P;
XX
DR   WPI; 2000-638211/61.
DR   P-PSDB; AAB34693.
XX
PT   Novel proteins and polypeptides useful for the treatment of e.g
PT   multiple sclerosis, systemic lupus erythmatosus, rheumatoid arthritis,
PT   cancer, Alzheimer's disease, Parkinson's disease, stroke, anemia and
PT   ulcers -
XX
PS   Claim 22; Page 394-395; 493pp; English.
XX
CC   This invention relates to 59 human secreted proteins and the nucleotide
CC   sequences encoding them. Sequences AAC59788-C59846 and AAB34687-B34745
CC   represent the proteins and their encoding nucleotide sequences, and
CC   sequences AAB34746-B34771 represent fragments of the proteins. Probes
CC   for the DNA sequences are represented by sequences AAC59847-C59596. The
```

```
CC   proteins exhibit neuroprotective, dermatological, immunosuppressive,
CC   antiinflammatory, antianaemic, nootropic, antiparkinsonian,
CC   cerebroprotective, haemostatic, vulnerary, cytostatic, antipsoriatic,
CC   antibacterial, virucide, and fungicide activity. The proteins and
CC   nucleotide sequences are useful as nutritional sources or supplements
CC   and in research. The proteins are useful for treating immune deficiency
CC   and disorders, which may be genetic or resulting from infections,
CC   autoimmune disorders such as multiple sclerosis, systemic lupus
CC   erythmatosus, rheumatoid arthritis, and for treating myeloid or lymphoid
CC   cell deficiencies such as anaemias by regulating haematopoiesis. The
CC   proteins are also useful in compositions for bone, cartilage, tendon,
CC   ligament and/or nerve tissue growth or regeneration, for wound healing,
CC   tissue repair and replacement and in the treatment of wounds, incisions
CC   and ulcers. Other uses include in the treatment of central and
CC   peripheral nervous system and neuropathies such as Alzheimer's and
CC   Parkinson's diseases and Shy-Drager syndrome, and mechanical and
CC   traumatic disorders, such as spinal cord disorders, head trauma and
CC   stroke. The proteins may also be used as a contraceptive, and for
CC   treating coagulation disorders such as haemophilias. The protein and
CC   nucleotide sequences with cadherin activity are useful for treating
CC   cancer. Other uses for the protein include for inhibiting the growth,
CC   infection or function of, or killing, infectious agents such as bacteria,
CC   virus, fungi and other parasites, for effecting bodily characteristics
CC   such as height, weight, hair colour, effecting biorhythms or cardiac
CC   cycles or rhythms, effecting metabolism, catabolism, anabolism,
CC   processing, utilization, storage or elimination of dietary fat, lipid,
CC   protein, carbohydrate, vitamins, minerals, cofactors, effecting
CC   behavioural characteristics, providing analgesic effects and for treating
CC   hyperproliferative disorders such as psoriasis.
XX
SQ   Sequence 1718 BP; 297 A; 579 C; 535 G; 307 T; 0 other;

  Query Match             98.5%;  Score 1088.2;  DB 21;  Length 1718;
  Best Local Similarity   99.6%;  Pred. No. 4.6e-194;
  Matches 1101;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy      1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    145 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 204

Qy     61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    205 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 264

Qy    121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
          |||||||| |||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    265 tgttccgccggagccccggccaaatacagcatcaccttcacgggcaagtggagccagacg 324

Qy    181 gccttccccaagcagtacccccctgttccgccccccctgcccagtggtcttcgctgctgggg 240
          | |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    325 g-cttccccaagcagtacccccctgttccgccccccctgcgcagtggtcttcgctgctgggg 383

Qy    241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    384 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 443

Qy    301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    444 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 503

Qy    361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    504 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 563

Qy    421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    564 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 623

Qy    481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    624 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 683

Qy    541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    684 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 743

Qy    601 ttcaccttctcctccccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    744 ttcaccttctcctccccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 803

Qy    661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    804 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 863

Qy    721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    864 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 923

Qy    781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    924 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 983

Qy    841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    984 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 1043

Qy    901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1044 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 1103

Qy    961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1104 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1163

Qy   1021 agccccgcagcccctgggcccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1164 agccccgcagcccctgggcccccccggagccatggggtgtcgggggctcctgtgcaggct 1223

Qy   1081 catgctgcaggcggccgagggcaca 1105
          ||||||||||||||||||||||||
Db   1224 catgctgcaggcggccgagggcaca 1248

RESULT  15
AAV72537
ID   AAV72537 standard; cDNA; 996 BP.
XX
AC   AAV72537;
XX
DT   01-MAR-1999  (first entry)
XX
DE   Human mindin cDNA.
XX
KW   Mindin; human; integrin ligand; angiogenic disease; cancer;
KW   metastasis; chronic inflammatory disorder; rheumatoid arthritis;
KW   atherosclerosis; macular degeneration, diabetic retinopathy;
KW   restenosis; Alzheimer's disease; neural disorder;
KW   tissue remodelling; therapy; diagnosis; ss.
XX
OS   Homo sapiens.
XX
PN   WO9850073-A1.
XX
PD   12-NOV-1998.
XX
PF   07-MAY-1998;   98WO-US09476.
XX
PR   09-MAY-1997;   97US-0046106.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Jonak ZL,  Lane PA,  Trulli SH,  Tsui P;
XX
DR   WPI; 1999-034688/03.
DR   P-PSDB; AAW83328.
```

XX
PT A new polypeptide is the integrin ligand, human mindin - useful in
PT therapy and diagnostic assays for diseases such as those associated
PT with angiogenesis
XX
PS Claim 2; Page 28; 39pp; English.
XX
CC This cDNA clone codes for human mindin (see AAW83328), a novel
CC integrin ligand. It has homology to F spondin, and may have a
CC multifunctional activity in normal and disease states. The
CC invention provides human mindin polypeptides and polynucleotides
CC and methods for producing such polypeptides by recombinant
CC techniques. It also relates to methods for using such polypeptides
CC and polynucleotides in the treatment of angiogenic diseases (cancer,
CC cancer metastasis, chronic inflammatory disorders, rheumatoid
CC arthritis, atherosclerosis, macular degeneration, diabetic
CC retinopathy), restenosis, Alzheimer's disease, neural disorders and
CC tissue remodelling. The invention also relates to methods for
CC identifying agonists and antagonists/inhibitors and for treating
CC conditions associated with human mindin imbalance with the
CC identified compounds. Diagnostic assays for detecting diseases
CC associated with inappropriate human mindin activity or levels are
CC also claimed.
XX
SQ Sequence 996 BP; 167 A; 352 C; 317 G; 160 T; 0 other;

Query Match 89.6%; Score 989.6; DB 20; Length 996;
Best Local Similarity 99.6%; Pred. No. 9.7e-176;
Matches 992; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy    19 atggaaaaccccagcccggccgccgccctgggcaaggccctctgcgctctcctcctggcc 78
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 atggaaaaccccagcccggccgccgccctgggcaaggccctctgcgctctcctcctggcc 60

Qy    79 actctcggcgccgccggccagcctcttgggggagagtccatctgttccgccagagccctg 138
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 actctcggcgccgccggccagcctcttgggggagagtccatctgttccgccagagccctg 120

Qy   139 gccaaatacagcatcaccttcacgggcaagtggagccagacggccttccccaagcagtac 198
         |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   121 gccaaatacagcatcaccttcacgggcaattggagccagacggccttccccaagcagtac 180

Qy   199 cccctgttccgcccccctgcccagtggtcttcgctgctgggggccgcgcatagctccgac 258
         |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   181 cccctgttccgcccccctgcgcagtggtcttcgctgctgggggccgcgcatagctccgac 240

Qy   259 tacagcatgtggaggaagaaccagtacgtcagtaacgggctgcgcgactttgcggagcgc 318
         |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   241 tacagcatgtggaggaagaaccagtacgtcagtaacgggctgcgcgactttgcggaacgc 300

Qy   319 ggcgaggcctgggcgctgatgaaggagatcgaggcggcgggggaggcgctgcagagcgtg 378
         |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   301 ggcgaggcctgggcgctgattaaggagatcgaggcggcgggggaggcgctgcagagcgtg 360

Qy   379 cacgcggtgttttcggcgcccgccgtccccagcggcaccgggcagacgtcggcggagctg 438
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 cacgcggtgttttcggcgcccgccgtccccagcggcaccgggcagacgtcggcggagctg 420

Qy   439 gaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgcatcgtgcccagccccgac 498
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 gaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgcatcgtgcccagccccgac 480

Qy   499 tggttcgtgggcgtggacagcctggacctgtgcgacggggaccgttggcgggaacaggcg 558
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 tggttcgtgggcgtggacagcctggacctgtgcgacggggaccgttggcgggaacaggcg 540

Qy   559 gcgctggacctgtacccctacgacgccgggacggacagcggcttcaccttctcctccccc 618
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 gcgctggacctgtacccctacgacgccgggacggacagcggcttcaccttctcctccccc 600

Qy   619 aacttcgccaccatcccgcaggacacggtgaccgagataacgtcctcctctcccagccac 678
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 aacttcgccaccatcccgcaggacacggtgaccgagataacgtcctcctctcccagccac 660

Qy   679 ccggccaactccttctactacccgcggctgaaggccctgcctcccatcgccagggtgaca 738
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ccggccaactccttctactacccgcggctgaaggccctgcctcccatcgccagggtgaca 720

Qy   739 ctggtgcggctgcgacagagccccagggccttcatccctcccgccccagtcctgcccagc 798
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 ctggtgcggctgcgacagagccccagggccttcatccctcccgccccagtcctgcccagc 780

Qy   799 agggacaatgagattgtagacagcgcctcagttccagaaacgccgctggactgcgaggtc 858
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 agggacaatgagattgtagacagcgcctcagttccagaaacgccgctggactgcgaggtc 840

Qy   859 tccctgtggtcgtcctggggactgtgcggaggccactgtgggaggctcgggaccaagagc 918
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 tccctgtggtcgtcctggggactgtgcggaggccactgtgggaggctcgggaccaagagc 900

Qy   919 aggactcgctacgtccgggtccagcccgccaacaacgggagcccctgccccgagctcgaa 978
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 aggactcgctacgtccgggtccagcccgccaacaacgggagcccctgccccgagctcgaa 960

Qy   979 gaagaggctgagtgcgtccctgataactgcgtctaa 1014
         ||||||||||||||||||||||||||||||||||||
Db   961 gaagaggctgagtgcgtccctgataactgcgtctaa 996
```

Search completed: December 10, 2001, 19:17:34
Job time: 20408 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2001, 13:37:26 ; Search time 89.78 Seconds
(without alignments)
2787.460 Million cell updates/sec

Title: US-09-170-042A-1
Perfect score: 1105
Sequence: 1 cgctgctcctgccgggtgat.....tgcaggcggccgagggcaca 1105

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1105	100.0	1105	2	US-08-799-173A-1	Sequence 1, Appli
	2	1051.8	95.2	1779	4	US-09-371-696-1	Sequence 1, Appli
	3	343.8	31.1	400	1	US-08-644-326-1	Sequence 1, Appli
	4	339	30.7	400	4	US-09-022-238-1	Sequence 1, Appli
	5	300	27.1	506	2	US-08-799-173A-15	Sequence 15, Appl
	6	202.6	18.3	316	2	US-08-799-173A-16	Sequence 16, Appl
	7	202.6	18.3	316	2	US-08-799-173A-17	Sequence 17, Appl
	8	70	6.3	4029	1	US-07-862-021B-9	Sequence 9, Appli
	9	70	6.3	4029	1	US-08-313-288B-9	Sequence 9, Appli
	10	70	6.3	4029	5	PCT-US93-03164-9	Sequence 9, Appli
	11	61.6	5.6	3226	1	US-07-862-021B-11	Sequence 11, Appl
	12	61.6	5.6	3226	1	US-08-313-288B-11	Sequence 11, Appl
	13	61.6	5.6	3226	5	PCT-US93-03164-11	Sequence 11, Appl
	14	57.2	5.2	1816	1	US-07-862-021B-13	Sequence 13, Appl
	15	57.2	5.2	1816	1	US-08-313-288B-13	Sequence 13, Appl
	16	57.2	5.2	1816	5	PCT-US93-03164-13	Sequence 13, Appl
c	17	57	5.2	1105	2	US-08-799-173A-1	Sequence 1, Appli
	18	56	5.1	933	4	US-09-105-390-43	Sequence 43, Appl
	19	56	5.1	1008	4	US-09-105-390-59	Sequence 59, Appl
	20	56	5.1	1155	4	US-08-818-112-12	Sequence 12, Appl
	21	56	5.1	2810	4	US-09-105-390-6	Sequence 6, Appli
	22	55	5.0	30001	1	US-08-125-468-1	Sequence 1, Appli
	23	55	5.0	30001	2	US-08-474-933-1	Sequence 1, Appli
c	24	54	4.9	1779	4	US-09-371-696-1	Sequence 1, Appli
	25	52.4	4.7	33529	4	US-09-144-085-3	Sequence 3, Appli
	26	51.8	4.7	2635	3	US-09-126-280-3	Sequence 3, Appli
	27	51.8	4.7	2670	3	US-09-126-280-1	Sequence 1, Appli
	28	50.2	4.5	2961	2	US-08-407-875-1	Sequence 1, Appli
	29	50.2	4.5	8625	4	US-08-980-832-1	Sequence 1, Appli
	30	50.2	4.5	11233	4	US-08-980-832-27	Sequence 27, Appl
	31	50	4.5	8051	2	US-08-576-626A-2	Sequence 2, Appli
c	32	49.8	4.5	1931	2	US-09-130-114-2	Sequence 2, Appli
c	33	49.2	4.5	1600	4	US-09-434-288-10	Sequence 10, Appl
	34	48.6	4.4	15664	1	US-08-402-282-3	Sequence 3, Appli
	35	48.6	4.4	15664	1	US-08-508-004-3	Sequence 3, Appli
	36	48.6	4.4	15664	1	US-08-402-066-3	Sequence 3, Appli
	37	48.6	4.4	15664	1	US-08-402-068-3	Sequence 3, Appli
	38	47.6	4.3	426	1	US-08-470-179-179	Sequence 179, App
c	39	47.4	4.3	5045	4	US-09-390-721-1	Sequence 1, Appli
c	40	47.4	4.3	5045	4	US-09-390-721-3	Sequence 3, Appli
	41	46.8	4.2	2353	5	PCT-US92-06840-1	Sequence 1, Appli
	42	46.8	4.2	43280	2	US-08-804-227C-1	Sequence 1, Appli
	43	46.2	4.2	3027	2	US-08-680-326-23	Sequence 23, Appl
	44	46	4.2	1281	4	US-09-105-537-19	Sequence 19, Appl
	45	46	4.2	13613	4	US-09-105-537-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-08-799-173A-1
; Sequence 1, Application US/08799173A
; Patent No. 5871969
; GENERAL INFORMATION:
;   APPLICANT: HASTINGS, GREGG,
;   APPLICANT: PATRICK J. DILLON
;   TITLE OF INVENTION: HUMAN NEURONAL ATTACHMENT FACTOR-1
;   NUMBER OF SEQUENCES: 18
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: HUMAN GENOME SCIENCES,INC.
;     STREET: 9410 KEY WEST AVENUE
;     CITY: ROCKVILLE
;     STATE: MD
;     COUNTRY: USA
;     ZIP: 20850
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/799,173A
;     FILING DATE: 11-FEB-1997
;     CLASSIFICATION: 536
;   ATTORNEY/AGENT INFORMATION:
;     NAME: BROOKES, ANDERS A.
;     REGISTRATION NUMBER: 36,373
;     REFERENCE/DOCKET NUMBER: PF226
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (301) 309-8504
;     TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1105 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 19..1011
;   FEATURE:
;     NAME/KEY: mat_peptide
;     LOCATION: 19..963
US-08-799-173A-1


  Query Match            100.0%;  Score 1105;  DB 2;  Length 1105;
  Best Local Similarity  100.0%;  Pred. No. 1.9e-200;
```

```
  Matches 1105;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGCTGCTCCTGCCGGGTGATGGAAAACCCCAGCCCGGCCGCCGCCCTGGGCAAGGCCCTC 60

Qy     61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGCGCTCTCCTCCTGGCCACTCTCGGCGCCGCCGGCCAGCCTCTTGGGGGAGAGTCCATC 120

Qy    121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGTTCCGCCAGAGCCCTGGCCAAATACAGCATCACCTTCACGGGCAAGTGGAGCCAGACG 180

Qy    181 gccttccccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GCCTTCCCCAAGCAGTACCCCCTGTTCCGCCCCCCTGCCCAGTGGTCTTCGCTGCTGGGG 240

Qy    241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCCGCGCATAGCTCCGACTACAGCATGTGGAGGAAGAACCAGTACGTCAGTAACGGGCTG 300

Qy    301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CGCGACTTTGCGGAGCGCGGCGAGGCCTGGGCGCTGATGAAGGAGATCGAGGCGGCGGGG 360

Qy    361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GAGGCGCTGCAGAGCGTGCACGCGGTGTTTTCGGCGCCCGCCGTCCCCAGCGGCACCGGG 420

Qy    421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CAGACGTCGGCGGAGCTGGAGGTGCAGCGCAGGCACTCGCTGGTCTCGTTTGTGGTGCGC 480

Qy    481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATCGTGCCCAGCCCCGACTGGTTCGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGGAC 540

Qy    541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CGTTGGCGGGAACAGGCGGCGCTGGACCTGTACCCCTACGACGCCGGGACGGACAGCGGC 600

Qy    601 ttcaccttctcctccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTCACCTTCTCCTCCCCCAACTTCGCCACCATCCCGCAGGACACGGTGACCGAGATAACG 660

Qy    661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TCCTCCTCTCCCAGCCACCCGGCCAACTCCTTCTACTACCCGCGGCTGAAGGCCCTGCCT 720

Qy    721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CCCATCGCCAGGGTGACACTGGTGCGGCTGCGACAGAGCCCCAGGGCCTTCATCCCTCCC 780

Qy    781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GCCCCAGTCCTGCCCAGCAGGGACAATGAGATTGTAGACAGCGCCTCAGTTCCAGAAACG 840

Qy    841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CCGCTGGACTGCGAGGTCTCCCTGTGGTCGTCCTGGGGACTGTGCGGAGGCCACTGTGGG 900

Qy    901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 AGGCTCGGGACCAAGAGCAGGACTCGCTACGTCCGGGTCCAGCCCGCCAACAACGGGAGC 960

Qy    961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CCCTGCCCCGAGCTCGAAGAAGAGGCTGAGTGCGTCCCTGATAACTGCGTCTAAGACCAG 1020

Qy   1021 agccccgcagcccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AGCCCCGCAGCCCCTGGGGCCCCCCGGAGCCATGGGGTGTCGGGGGCTCCTGTGCAGGCT 1080

Qy   1081 catgctgcaggcggccgagggcaca 1105
          |||||||||||||||||||||||||
Db   1081 CATGCTGCAGGCGGCCGAGGGCACA 1105


RESULT   2
US-09-371-696-1
; Sequence 1, Application US/09371696
; Patent No. 6287777
; GENERAL INFORMATION:
;  APPLICANT: Sytkowski, Arthur J.
;  APPLICANT: Yang, Meiheng
;  TITLE OF INVENTION: NOVEL NPG-1 GENE THAT IS DIFFERENTIALLY EXPRESSED IN PROSTATE
;  TITLE OF INVENTION: TUMORS
;  FILE REFERENCE: 01948/053002
;  CURRENT APPLICATION NUMBER: US/09/371,696
;  CURRENT FILING DATE: 1999-08-10
;  EARLIER APPLICATION NUMBER: US 09/022,238
;  EARLIER FILING DATE: 1998-02-11
;  EARLIER APPLICATION NUMBER: US 08/644,326
;  EARLIER FILING DATE: 1996-05-10
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 1779
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (236)...(1225)
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1779)
;   OTHER INFORMATION: n = A,T,C or G
US-09-371-696-1


  Query Match             95.2%;  Score 1051.8;  DB 4;  Length 1779;
  Best Local Similarity   98.2%;  Pred. No. 2.1e-190;
  Matches 1085;  Conservative    0;  Mismatches   17;  Indels    3;  Gaps    2;

Qy      1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    218 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 277

Qy     61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    278 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 337

Qy    121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    338 tgttccgccagagccccggccaaatacagcatcaccttcacgggcaagtggagccagacg 397

Qy    181 gccttccccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggg 240
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    398 gccttccccaagcagtacccccctgttccgcccccctgcgcagtggtcttcgctgctgggg 457

Qy    241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    458 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 517

Qy    301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    518 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 577

Qy    361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    578 gaggcgctgcagagcgtgcacgaggtgttttcggcgcccgccgtccccagcggcaccggg 637

Qy    421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    638 cagacgtcggcggagctggaggtgcagcgcaagcactcgctggtctcgtttgtggtgcgc 697
```

```
Qy     481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     698 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 757

Qy     541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     758 cgttggcgggaacaggcggcgcttgacctgtacccctacgacgccgggacggacagcggc 817

Qy     601 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     818 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 877

Qy     661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
           |||||||||||||||||||| |||||||||||||||||||||||    ||
Db     878 tcctcctctcccagccacccggccagctccttctactacccgcggctgaaggcc--tgct 935

Qy     721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     936 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 995

Qy     781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
           |||||||||||||||||| |||||||| | ||||||||||||||||||||||||||||||
Db     996 gccccagtcctgcccagca-ggacaatgcgcttgtagacagcgcctcagttccagaaaca 1054

Qy     841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1055 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 1114

Qy     901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
           ||||||||| ||||||||||||||  |||| ||||||||||||||||||||||||||||||
Db    1115 aggctcgggtccaagagcaggactccgtacgcccgggtccagcccgccaacaacgggagc 1174

Qy     961 ccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgcgtctaagaccag 1020
           |||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db    1175 ccctgccccgagctcgaagaagaggctgagtgcgtggctgataactgcgtctaagaccag 1234

Qy    1021 agccccgcagccccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1235 agccccgcagccccctggggccccccggagccatggggtgtcgggggctcctgtgcaggct 1294

Qy    1081 catgctgcaggcggccgagggcaca 1105
           ||||||||||||||||||||||||||
Db    1295 catgctgcaggcggccgagggcaca 1319


RESULT   3
US-08-644-326-1
; Sequence 1, Application US/08644326
; Patent No. 5804382
;  GENERAL INFORMATION:
;    APPLICANT:  Arthur J. Sytkowski and Meiheng Yang
;    TITLE OF INVENTION:  Methods for identifying Differentially Expressed Genes and di
;    TITLE OF INVENTION:  genomic nucleic acid sequences
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/644,326
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ralph A. Loren
;      REGISTRATION NUMBER:  29,325
;      REFERENCE/DOCKET NUMBER:  NER-262
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  400 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-644-326-1


  Query Match             31.1%;  Score 343.8;  DB 1;  Length 400;
  Best Local Similarity   97.0%;  Pred. No. 5.2e-57;
  Matches  393;  Conservative    0;  Mismatches    7;  Indels    5;  Gaps    4;

Qy     570 gtacccctacgacgccgggacggacagcggcttcaccttctcctcccccaacttcgccac 629
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1 GTACCCCTACGACGCCGGGAC-GACAGCGGCTTCACCTTCTCCTCCCCCAACTTCGCCAC 59

Qy     630 catcccgcaggacacggtgaccgagataacgtcctcctctcccagccacccggccaactc 689
           ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||| |||
Db      60 CATCCCGCA-GACACGGTGACCGAGATAACGTCCTCCTCTCCCAGCCACCCGGCCAGCTC 118

Qy     690 cttctactacccgcggctgaaggccctgcctcccatcgccagggtgacactggtgcggct 749
           ||||||||||||||||||||||||    ||||||||||||||||||||||||||||||||
Db     119 CTTCTACTACCCGCGGCTGAAGGCC--TGCTCCCATCGCCAGGGTGACACTGGTGCGGCT 176

Qy     750 gcgacagagccccagggccttcatccctcccgccccagtcctgcccagcagggacaatga 809
           |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     177 GCGACAGAGCCCCAGGGCCTTCATCCCTCCCGCCCCAGTCCTGCCCAGCA-GGACAATGA 235

Qy     810 gattgtagacagcgcctcagttccagaaacgccgctggactgcgaggtctccctgtggtc 869
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     236 GATTGTAGACAGCGCCTCAGTTCCAGAAACGCCGCTGGACTGCGAGGTCTCCCTGTGGTC 295

Qy     870 gtcctggggactgtgcggaggccactgtgggaggctcgggaccaagagcaggactcgcta 929
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||  ||
Db     296 GTCCTGGGGACTGTGCGGAGGCCACTGTGGGAGGCTCGGGTCCAAGAGCAGGACTCCGTA 355

Qy     930 cgtccgggtccagcccgccaacaacgggagcccctgccccgagct 974
           || ||||||||||||||||||||||||||||||||||||||||||
Db     356 CGCCCGGGTCCAGCCCGCCAACAACGGGAGCCCCTGCCCCGAGCT 400


RESULT   4
US-09-022-238-1
; Sequence 1, Application US/09022238
; Patent No. 6177244
;  GENERAL INFORMATION:
;    APPLICANT:  Sytkowski, Arthur J. and Yang, Meiheng
;    TITLE OF INVENTION:  A novel NPG-1 Gene that is differentially expressed in pros
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/022,238
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/644,326
;      FILING DATE:  10-MAY-1996
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maravic-Magovcevic, Ivana
;      REGISTRATION NUMBER:  P-43,338
;      REFERENCE/DOCKET NUMBER:  NER-262CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)742-4214
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  400 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..398
US-09-022-238-1


 Query Match             30.7%;  Score 339;  DB 4;  Length 400;
 Best Local Similarity   96.3%;  Pred. No. 4.2e-56;
 Matches  390;  Conservative    0;  Mismatches   10;  Indels    5;  Gaps    4;

Qy    570 gtacccctacgacgccgggacggacagcggcttcaccttctcctccccccaacttcgccac 629
          ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db      1 GTACCCCTACGACGCCGGGAC-GACAGCGGCTTCACCTTCTCCTCCCCCAACTTCGCCAC 59

Qy    630 catcccgcaggacacggtgaccgagataacgtcctcctctcccagccacccggccaactc 689
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||| |||
Db     60 CATCCCGCA-GACACGGTGACCGAGATAACGTCCTCCTCTCCCAGCCACCCGGCCAGCTC 118

Qy    690 cttctactacccgcggctgaaggccctgcctcccatcgccagggtgacactggtgcggct 749
          ||||||||||||||||||||||    |||||||||||||||||||||||||||||||||
Db    119 CTTCTACTACCCGCGGCTGAAGGCC--TGCTCCCATCGCCAGGGTGACACTGGTGCGGCT 176

Qy    750 gcgacagagccccagggccttcatccctcccgccccagtcctgcccagcagggacaatga 809
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    177 GCGACAGAGCCCCAGGGCCTTCATCCCTCCCGCCCCAGTCCTGCCCAGCA-GGACAATGC 235

Qy    810 gattgtagacagcgcctcagttccagaaacgccgctggactgcgaggtctccctgtggtc 869
          | |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db    236 GCTTGTAGACAGCGCCTCAGTTCCAGAAACACCGCTGGACTGCGAGGTCTCCCTGTGGTC 295

Qy    870 gtcctggggactgtgcggaggccactgtgggaggctcgggaccaagagcaggactcgcta 929
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||  ||
Db    296 GTCCTGGGGACTGTGCGGAGGCCACTGTGGGAGGCTCGGGTCCAAGAGCAGGACTCCGTA 355

Qy    930 cgtccgggtccagcccgccaacaacgggagcccctgccccgagct 974
          || |||||||||||||||||||||||||||||||||||||||||
Db    356 CGCCCGGGTCCAGCCCGCCAACAACGGGAGCCCCTGCCCCGAGCT 400


RESULT   5
US-08-799-173A-15
; Sequence 15, Application US/08799173A
; Patent No. 5871969
;  GENERAL INFORMATION:
;    APPLICANT:  HASTINGS, GREGG,
;    APPLICANT:  PATRICK J. DILLON
;    TITLE OF INVENTION:  HUMAN NEURONAL ATTACHMENT FACTOR-1
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HUMAN GENOME SCIENCES,INC.
;      STREET:  9410 KEY WEST AVENUE
;      CITY:  ROCKVILLE
;      STATE:  MD
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/799,173A
;      FILING DATE:  11-FEB-1997
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROOKES, ANDERS A.
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PF226
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  506 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-799-173A-15


  Query Match             27.1%;  Score 300;  DB 2;  Length 506;
  Best Local Similarity   84.8%;  Pred. No. 9.5e-49;
  Matches  368;  Conservative    0;  Mismatches   62;  Indels    4;  Gaps    4;

Qy     13 cgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctctgcgctctcctc 72
          |||   | | |||||||||||||||| ||||||||||||||||  ||||||||||||
Db      6 CGGCANAGGNNAAACCCCAGCCCGGCTGCCGCCCTGGGCAAGGCCTNCTGCGCTCTCCTC 65

Qy     73 ctggccactctcggcgccgccggccagcctcttgggggagagtccatctgttccgccaga 132
          |||||||||||||||||| |  |||||||||||||||||||||||||| ||||||||||
Db     66 CTGGCCACTCTCGGCGCCGGC-ACCAGCCTCTTGGGGGAGAGTCCATCTNTTCCGCCAGA 124

Qy    133 gccctggccaaatacagcatcaccttcacgggcaagtggagccagacggccttccccaag 192
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    125 GCCCCGGCCAAATACAGCATCACCTTCACGGGCAAGTGGAGCCAGACGGCCTTCCCCAAG 184

Qy    193 cagtacccccctgttccgccccccctgcccagtggtcttcgctgctgggggccgcgcatagc 252
          |||||||||||||||||||||||||| || |||| ||||||||||||||||||||||||
Db    185 CAGTACCCCCCTGTTCCGCCCCCCCTGCGCA-TGGTNTTCGCTGCTGGGGGCCGCGCATAGC 243

Qy    253 tccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctgcgcgactttgcg 312
          |||||||||||||||||||||||||||||||||||||  ||||||||||||||||| ||
Db    244 TCCGACTACAGCATGTGGAGGAAGAACCAGTACGTCATAAACGGGCTGCGCGACTTTNCG 303

Qy    313 gagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggggaggcgctgcag 372
          |||||  | | | |  || || ||||||||||||||  || ||||||||||||    |
Db    304 GAGCGGCGAG-GCCTNGGNCGTTGATGAAGGAGATCCGGGNGGCGGGGGAGGCGTNCAAN 362

Qy    373 agcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccgggcagacgtcggcg 432
          ||    || | | | ||||||| |||||   ||||||  || | | || | |||| || |
Db    363 AGGTGNCAAGAGTTNTTTTCGGGGCCCG-GTTCCCCAANGGNAACNGGNAAACGTTGGGG 421

Qy    433 gagctggaggtgca 446
          | |  || | |
Db    422 GNTTTNNAGTTTNA 435


RESULT   6
US-08-799-173A-16
; Sequence 16, Application US/08799173A
; Patent No. 5871969
;  GENERAL INFORMATION:
;    APPLICANT:  HASTINGS, GREGG,
;    APPLICANT:  PATRICK J. DILLON
;    TITLE OF INVENTION:  HUMAN NEURONAL ATTACHMENT FACTOR-1
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HUMAN GENOME SCIENCES,INC.
;      STREET:  9410 KEY WEST AVENUE
;      CITY:  ROCKVILLE
;      STATE:  MD
```

```
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/799,173A
;      FILING DATE:  11-FEB-1997
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROOKES, ANDERS A.
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PF226
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  316 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-799-173A-16


  Query Match             18.3%;  Score 202.6;  DB 2;  Length 316;
  Best Local Similarity   85.6%;  Pred. No. 2.1e-30;
  Matches  274;  Conservative    0;  Mismatches   39;  Indels    7;  Gaps    5;

Qy     272 ggaagaaccagtacgtcagtaacgggctgcgcgactttgcggagcgcggcgaggcctggg 331
           || |   |||||||||| ||||||||||||||||||||||  ||||||||||||||||
Db       1 GGCANNGCCAGTACGTCA-TAACGGGCTGCGCGACTTTGCGGA-NGCGGCGAGGCCTGGG 58

Qy     332 cgctgatgaaggagatcgaggcggcgggggaggcgctgcagagcgtgcacgcggtgtttt 391
           ||||||||||||||||| ||||||||||||||||||||||| ||||||| ||||||||
Db      59 CGCTGATGAAGGAGATCAAGGCGGCGGGGGAGGCGCTGCAGAG-GTGCACGAGGTGTTTT 117

Qy     392 cggcgcccgccgtccccagcggcaccgggcagacgtcggcggagctggaggtgcagcgca 451
           |||||||||     ||||||| |||| |||||||||||||| |   | ||||||||||||
Db     118 CGGCGCCCG-GTNNCCCAGCGNCACCNGGCAGACGTCGGCGAACTGGNAGGTGCAGCGCA 176

Qy     452 ggcactcgctggtctcgtttgtggtgcgcatcgtgcccagccccgactggttcgtgggcg 511
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     177 GGCACTCGCTGGTCTCGTTTGTGGTGCGCATCGTGCCCAGCCCCGACTGGTTCGTGGGCG 236

Qy     512 tggacagcctggacc---tgtgcgacggggaccgttggcgggaacaggcggcgctggacc 568
           ||||||||||||  |   ||      |||  | | |   | ||||||| || ||||||
Db     237 TGGACAGCCTGGGACCTGTGANAACGGGGACCTTTNGCGNGNAACAGGCGNCGTTGGACC 296

Qy     569 tgtacccctacgacgccggg 588
           |||| |||||||||| ||||
Db     297 TGTANCCCTACGACGNCGGG 316


RESULT   7
US-08-799-173A-17
; Sequence 17, Application US/08799173A
; Patent No. 5871969
;  GENERAL INFORMATION:
;    APPLICANT:  HASTINGS, GREGG,
;    APPLICANT:  PATRICK J. DILLON
;    TITLE OF INVENTION:  HUMAN NEURONAL ATTACHMENT FACTOR-1
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HUMAN GENOME SCIENCES,INC.
;      STREET:  9410 KEY WEST AVENUE
;      CITY:  ROCKVILLE
;      STATE:  MD
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/799,173A
;      FILING DATE:  11-FEB-1997
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROOKES, ANDERS A.
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PF226
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  316 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-799-173A-17


  Query Match             18.3%;  Score 202.6;  DB 2;  Length 316;
  Best Local Similarity   85.6%;  Pred. No. 2.1e-30;
  Matches  274;  Conservative    0;  Mismatches   39;  Indels    7;  Gaps    5;

Qy     272 ggaagaaccagtacgtcagtaacgggctgcgcgactttgcggagcgcggcgaggcctggg 331
           || |   |||||||||| ||||||||||||||||||||||  ||||||||||||||||
Db       1 GGCANNGCCAGTACGTCA-TAACGGGCTGCGCGACTTTGCGGA-NGCGGCGAGGCCTGGG 58

Qy     332 cgctgatgaaggagatcgaggcggcgggggaggcgctgcagagcgtgcacgcggtgtttt 391
           ||||||||||||||||| ||||||||||||||||||||||| ||||||| ||||||||
Db      59 CGCTGATGAAGGAGATCAAGGCGGCGGGGGAGGCGCTGCAGAG-GTGCACGAGGTGTTTT 117

Qy     392 cggcgcccgccgtccccagcggcaccgggcagacgtcggcggagctggaggtgcagcgca 451
           |||||||||     ||||||| |||| |||||||||||||| |   | ||||||||||||
Db     118 CGGCGCCCG-GTNNCCCAGCGNCACCNGGCAGACGTCGGCGAACTGGNAGGTGCAGCGCA 176

Qy     452 ggcactcgctggtctcgtttgtggtgcgcatcgtgcccagccccgactggttcgtgggcg 511
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     177 GGCACTCGCTGGTCTCGTTTGTGGTGCGCATCGTGCCCAGCCCCGACTGGTTCGTGGGCG 236

Qy     512 tggacagcctggacc---tgtgcgacggggaccgttggcgggaacaggcggcgctggacc 568
           ||||||||||||  |   ||      |||  | | |   | ||||||| || ||||||
Db     237 TGGACAGCCTGGGACCTGTGANAACGGGGACCTTTNGCGNGNAACAGGCGNCGTTGGACC 296

Qy     569 tgtacccctacgacgccggg 588
           |||| |||||||||| ||||
Db     297 TGTANCCCTACGACGNCGGG 316


RESULT   8
US-07-862-021B-9
; Sequence 9, Application US/07862021B
; Patent No. 5279966
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M
;    APPLICANT:  Klar, Avihu
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/862,021B
;      FILING DATE:  19920405
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4029 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  226..2647
US-07-862-021B-9


  Query Match             6.3%;  Score 70;  DB 1;  Length 4029;
  Best Local Similarity   50.0%;  Pred. No. 2.4e-05;
  Matches  268;  Conservative    0;  Mismatches  250;  Indels   18;  Gaps    3;

Qy    120 ctgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagac 179
          ||| | |||| | |     ||||| |||||  |||| ||    || || |||    |||
Db    819 CTGCTGCGCCTGCGGAACTGCCAAGTACAGACTCACGTTTTATGGGAACTGGTCGGAGAA 878

Qy    180 ggccttccccaagcagtacccccctgttccgccccccctgcccagtggtcttcgctgctggg 239
          | |    || ||| | |||||       || |   ||   || |||||| |  |  | ||
Db    879 GACTCATCCAAAGGATTACCC------TCGTCGGGCTAATCACTGGTCTGCCATCATTGG 932

Qy    240 ggccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggct 299
          |   | ||    |    |||||    |||||  | |      ||| | |||| | ||| |
Db    933 CGGATCCCACTCCAAGAACTACGTGCTGTGGGAGTACGGAGGGTATGCCAGTGAAGGGGT 992

Qy    300 gcgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcggg 359
          |  |||| || |  |||    |      ||  | | || ||        | |
Db    993 CAAGCAAGTTGCTGAACTTGGCTCACCAGTAAAAATGGAGGAAGAAATTCGACAACAGAG 1052

Qy    360 ggaggcgctgcagagcgtgcac---------gcggtgttttcggcgcccgccgtccccag 410
          ||  |   | |  |  ||   |         ||   ||   |  |             ||
Db   1053 TGATGAAGTCCTCACTGTCATCAAAGCCAAAGCCCAGTGGCCATCCTGGCAGCCTGTCAA 1112

Qy    411 cggcaccgggcagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtt 470
          |  |  |       | || || ||  |    ||| |  | |  | |    || | || ||
Db   1113 TGTGAGAGCAGCACCCTCAGCCGAATTCTCAGTGGACAGGACACGCCACTTGATGTCCTT 1172

Qy    471 tgtggtgcgcatcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtg 530
          |       || |  ||||| || ||||||  |||||||  |          ||| |||||
Db   1173 CCTAACCATGATGGGCCCCAGTCCTGACTGGAACGTGGGCCTATCTGCAGAGGATCTGTG 1232

Qy    531 cgacggggaccgttggcgggaacag---gcggcgctggacctgtacccctacgacgccgg 587
          |  |  |||  || |  |||  |||   | || || ||||||    ||||  || || ||
Db   1233 CACCAAGGAGTGTGGCTGGGTCCAGAAAGTGGTGCAGGACCTAATTCCCTGGGATGCTGG 1292

Qy    588 gacggacagcggcttcaccttctcctccccaacttcgccaccatcccgcaggaca 643
          ||||||||||||  | |||| |   || || |||    |||| || || |||||  |
Db   1293 CACGGACAGCGGGGTGACCTACGAGTCACCAAACAAGCCCACAATTCCTCAGGAAA 1348


RESULT   9
US-08-313-288B-9
; Sequence 9, Application US/08313288B
```

```
; Patent No. 5750502
; GENERAL INFORMATION:
;   APPLICANT:  Jessell, Thomas M. and Avihu Klar
;   TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;   TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;   NUMBER OF SEQUENCES:  20
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Cooper & Dunham LLP
;     STREET:  1185 Avenue of the Americas
;     CITY:  New York
;     STATE:  New York
;     COUNTRY:  USA
;     ZIP:  10036
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/313,288B
;     FILING DATE:  January 5, 1995
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  White, John P.
;     REGISTRATION NUMBER:  28,678
;     REFERENCE/DOCKET NUMBER:  40028-A-PCT-US
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (212) 278-0400
;     TELEFAX:  (212) 391-0526
;     TELEX:
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4029 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  226..2647
US-08-313-288B-9


  Query Match             6.3%;  Score 70;  DB 1;  Length 4029;
  Best Local Similarity   50.0%;  Pred. No. 2.4e-05;
  Matches  268;  Conservative    0;  Mismatches  250;  Indels   18;  Gaps    3;

Qy    120 ctgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagac 179
          ||| | |||| | |     ||||| |||||  |||| ||    || || |||    |||
Db    819 CTGCTGCGCCTGCGGAACTGCCAAGTACAGACTCACGTTTTATGGGAACTGGTCGGAGAA 878

Qy    180 ggccttccccaagcagtacccccctgttccgccccccctgcccagtggtcttcgctgctggg 239
          | |    || ||| | |||||       || |   ||   || |||||| |  |  | ||
Db    879 GACTCATCCAAAGGATTACCC------TCGTCGGGCTAATCACTGGTCTGCCATCATTGG 932

Qy    240 ggccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggct 299
          |   | ||    |    |||||    |||||  | |      ||| | |||| | ||| |
Db    933 CGGATCCCACTCCAAGAACTACGTGCTGTGGGAGTACGGAGGGTATGCCAGTGAAGGGGT 992

Qy    300 gcgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcggg 359
          |  |||| || |  |||    |      ||  | | || ||        | |
Db    993 CAAGCAAGTTGCTGAACTTGGCTCACCAGTAAAAATGGAGGAAGAAATTCGACAACAGAG 1052

Qy    360 ggaggcgctgcagagcgtgcac---------gcggtgttttcggcgcccgccgtccccag 410
          ||  |   | |  |  ||   |         ||   ||   |  |             ||
Db   1053 TGATGAAGTCCTCACTGTCATCAAAGCCAAAGCCCAGTGGCCATCCTGGCAGCCTGTCAA 1112

Qy    411 cggcaccgggcagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtt 470
          |  |  |       | || || ||  |    ||| |  | |  | |    || | || ||
Db   1113 TGTGAGAGCAGCACCCTCAGCCGAATTCTCAGTGGACAGGACACGCCACTTGATGTCCTT 1172

Qy    471 tgtggtgcgcatcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtg 530
          |       || |  ||||| || ||||||  |||||||  |          ||| |||||
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2001, 13:37:25 ; Search time 1541.1 Seconds
(without alignments)
7704.941 Million cell updates/sec

Title: US-09-170-042A-1
Perfect score: 1105
Sequence: 1 cgctgctcctgccgggtgat.....tgcaggcggccgagggcaca 1105

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	946	85.6	963	10	AL546511	AL546511 AL546511
	2	868.4	78.6	934	10	AL552365	AL552365 AL552365
	3	861.2	77.9	1009	10	AL554923	AL554923 AL554923
	4	837.2	75.8	904	10	AL551401	AL551401 AL551401
	5	826.8	74.8	979	10	AL551786	AL551786 AL551786
	6	762.6	69.0	801	10	AL554806	AL554806 AL554806
	7	701.6	63.5	825	10	AL543187	AL543187 AL543187
	8	685.2	62.0	971	10	AL550317	AL550317 AL550317
	9	653.6	59.1	931	10	AL547912	AL547912 AL547912
	10	621.4	56.2	935	10	AL532820	AL532820 AL532820
c	11	613.8	55.5	989	10	AL564981	AL564981 AL564981
	12	600.6	54.4	838	10	AL551123	AL551123 AL551123
	13	600.4	54.3	682	10	BE407510	BE407510 601300346
c	14	593	53.7	1025	10	AL574221	AL574221 AL574221
	15	579.2	52.4	904	10	AL542564	AL542564 AL542564
	16	518.4	46.9	843	11	BG741018	BG741018 602634745
c	17	511	46.2	1011	10	AL574979	AL574979 AL574979
c	18	494.4	44.7	987	10	AL571835	AL571835 AL571835
c	19	492.4	44.6	981	10	AL575679	AL575679 AL575679
c	20	491.8	44.5	950	10	AL577524	AL577524 AL577524
	21	479	43.3	479	10	BE047605	BE047605 tz39c09.y
c	22	477	43.2	804	10	AL572678	AL572678 AL572678
	23	464	42.0	1046	11	BI411153	BI411153 602965629
c	24	460.2	41.6	831	10	AL542563	AL542563 AL542563
c	25	459.4	41.6	741	10	AL575359	AL575359 AL575359
c	26	446.4	40.4	854	10	AL577445	AL577445 AL577445
c	27	407.6	36.9	865	10	AL569820	AL569820 AL569820
	28	377	34.1	955	11	BF577395	BF577395 602092222
	29	372.6	33.7	805	11	BF161377	BF161377 601770148
	30	356.6	32.3	553	10	AW140484	AW140484 EST290465
	31	347.4	31.4	870	11	BI110293	BI110293 602900723
	32	334.2	30.2	516	10	AA801239	AA801239 EST190736
c	33	293	26.5	629	10	AL574780	AL574780 AL574780
	34	284.6	25.8	922	10	AL514050	AL514050 AL514050
	35	279.2	25.3	377	10	AA047125	AA047125 zk79b09.r
	36	268.2	24.3	413	10	AW142184	AW142184 EST292416
	37	262.2	23.7	712	11	BG865607	BG865607 602783761
	38	250.6	22.7	384	11	BF556916	BF556916 UI-R-A1-e
	39	246.8	22.3	516	10	BE689392	BE689392 uw54b01.y
	40	243.4	22.0	628	11	BG871696	BG871696 602791920
	41	243.4	22.0	698	11	BG174990	BG174990 602335886
	42	241.2	21.8	1137	12	AK009820	AK009820 Mus muscu
	43	233.4	21.1	432	10	AA671880	AA671880 v107d08.r
	44	232.8	21.1	362	10	AI120329	AI120329 ub69d02.r
c	45	219	19.8	632	11	BF475565	BF475565 nac46d12.

ALIGNMENTS

RESULT 1
AL546511
LOCUS AL546511 963 bp mRNA EST 16-FEB-2001
DEFINITION AL546511 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI030YP14 5 prime, mRNA sequence.
ACCESSION AL546511
VERSION AL546511.1 GI:12879698
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 963)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES Location/Qualifiers
source 1. .963
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DI030YP14"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 153 a 345 c 312 g 151 t 2 others
ORIGIN

Query Match 85.6%; Score 946; DB 10; Length 963;
Best Local Similarity 99.5%; Pred. No. 2.5e-158;
Matches 957; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

```
Qy       1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3 CGCTGCTCCTGCCGGGTGATGGAAAACCCCAGCCCGGCCGCCGCCCTGGGCAAGGCCCTC 62

Qy      61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63 TGCGCTCTCCTCCTGGCCACTCTCGGCGCCGCCGGCCAGCCTCTTGGGGGAGAGTCCATC 122

Qy     121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     123 TGTTCCGCCAGAGCCCCGGCCAAATACAGCATCACCTTCACGGGCAAGTGGAGCCAGACG 182

Qy     181 gccttccccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggg 240
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     183 GCCTTCCCCAAGCAGTACCCCCTGTTCCGCCCCCCTGCGCAGTGGTCTTCGCTGCTGGGG 242

Qy     241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     243 GCCGCGCATAGCTCCGACTACAGCATGTGGAGGAAGAACCAGTACGTCAGTAACGGGCTG 302

Qy     301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     303 CGCGACTTTGCGGAGCGCGGCGAGGCCTGGGCGCTGATGAAGGAGATCGAGGCGGCGGGG 362

Qy     361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     363 GAGGCGCTGCAGAGCGTGCACGCGGTGTTTTCGGCGCCCGCCGTCCCCAGCGGCACCGGG 422

Qy     421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     423 CAGACGTCGGCGGAGCTGGAGGTGCAGCGCAGGCACTCGCTGGTCTCGTTTGTGGTGCGC 482

Qy     481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     483 ATCGTGCCCAGCCCCGACTGGTTCGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGGAC 542

Qy     541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     543 CGTTGGCGGGAACAGGCGGCGCTGGACCTGTACCCCTACGACGCCGGGACGGACAGCGGC 602

Qy     601 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     603 TTCACCTTCTCCTCCCCCAACTTCGCCACCATCCCGCAGGACACGGTGACCGAGATAACG 662

Qy     661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     663 TCCTCCTCTCCCAGCCACCCGGCCAACTCCTTCTACTACCCGCGGCTGAAGGCCCTGCCT 722

Qy     721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     723 CCCATCGCCAGGGTGACACTGGTGCGGCTGCGACAGAGCCCCAGGGCCTTCATCCCTCCC 782

Qy     781 gccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacg 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     783 GCCCCAGTCCTGCCCAGCAGGGACAATGAGATTGTAGACAGCGCCTCAGTTCCAGAAACG 842

Qy     841 ccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtggg 900
           ||||||||||||||||||||||||||||||||||||||||||:|| |||||||||
Db     843 CCGCTGGACTGCGAGGTCTCCCTGTGGTCGTCCTGGGGACTGTGCGGRGG-CACTGTGGG 901

Qy     901 aggctcgggaccaagagcaggactcgctacgtccgggtccagcccgccaacaacgggagc 960
           ||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db     902 AGGCTCGGGACCAAGAGCAGGACTCGCTACGTCCGGGTCCAGCCCGCCAAMAACGGGAGC 961
```

Qy 961 cc 962
||
Db 962 CC 963

RESULT 2
AL552365
LOCUS AL552365 934 bp mRNA EST 16-FEB-2001
DEFINITION AL552365 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI069YF21 5 prime, mRNA sequence.
ACCESSION AL552365
VERSION AL552365.1 GI:12891195
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 934)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES Location/Qualifiers
source 1. .934
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DI069YF21"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT 149 a 327 c 303 g 151 t 4 others
ORIGIN

Query Match 78.6%; Score 868.4; DB 10; Length 934;
Best Local Similarity 98.6%; Pred. No. 1.4e-144;
Matches 924; Conservative 3; Mismatches 5; Indels 5; Gaps 5;

```
Qy       3 ctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctctg 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTGCTCCTGCCGGGTGATGGAAAACCCCAGCCCGGCCGCCGCCCTGGGCAAGGCCCTCTG 60

Qy      63 cgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatctg 122
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db      61 CGCTCTCCTCCTGGCCACTCTCGGCGCCGCCGGCCAGCCTCTTGGGGGAGAATCCATCTG 120

Qy     123 ttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacggc 182
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     121 TTCCGCCAGAGCCCCGGCCAAATACAGCATCACCTTCACGGGCAAGTGGAGCCAGACGGC 180

Qy     183 cttccccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggggc 242
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     181 CTTCCCCAAGCAGTACCCCCTGTTCCGCCCCCCTGCGCAGTGGTCTTCGCTGCTGGGGGC 240

Qy     243 cgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctgcg 302
           ||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db     241 CGCGCATAGCTCCGACTACAGCATGTGGAGGAAGAACCAGTACGTCAGTWACGGGCTGCG 300

Qy     303 cgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggggga 362
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    301 CGACTTTGCGGAG-GCGGCGAGGCCTGGGCGCTGATGAAGGAGATCGAGGCGGCGGGGGA 359

Qy    363 ggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccgggca 422
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    360 GGCGCTGCAGAGCGTGCACGCGNTGTTTTCGGCGCCCGCCGTCCCCAGCGGCACCGGGCA 419

Qy    423 gacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgcat 482
          |||||:|||||| :|||||||||||||||||||||||||||||||||||||||||||||||
Db    420 GACGTYGGCGGA-STGGAGGTGCAGCGCAGGCACTCGCTGGTCTCGTTTGTGGTGCGCAT 478

Qy    483 cgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggaccg 542
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    479 CGTGCCCAGCCCCGACTGGTTCGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGGACCG 538

Qy    543 ttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggctt 602
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    539 TTGGCGGGAACAGGCGGCGCTGGACCTGTACCCCTACGACGCCGGGACGGACAGCGGCTT 598

Qy    603 caccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacgtc 662
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    599 CACCTTCTCCTCCCCCAACTTCGCCACCATCCCGCA-GACACGGTGACCGAGATAACGTC 657

Qy    663 ctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcctcc 722
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    658 CTCCTCTCCCAGCCACCCGGCCAACTCCTTCTACTACCCGCGGCTGAAGGCCCTGCCTCC 717

Qy    723 catcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctcccgc 782
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    718 CATCGCCAGGGTGACACTGGTGCGGCTGCGACAGAGCCCCAGGGCCTTCATCCCTCCCGC 777

Qy    783 cccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacgcc 842
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    778 CCCAGTCCTGCCCAGCAGGGACAATGAGATTGTAGACAGCGCCTCAGTTCCAGAAACGCC 837

Qy    843 gctggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtgggag 902
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   -838 GCTGGACTGCGAGGTCTCCCTGTGGTCGTCCTGGGGACTGTGCGGAGG-CACTGTGGGAG 896

Qy    903 gctcgggaccaagagcaggactcgctacgtccgggtc 939
          |||||||| |||||||||||||||||| |||||
Db    897 GCTCGGGA-CAAGAGCAGGACTCGCTACGTCGGGGTC 932


RESULT   3
AL554923
LOCUS       AL554923     1009 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL554923 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI087YL11 5
            prime, mRNA sequence.
ACCESSION   AL554923
VERSION     AL554923.1  GI:12896167
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1009)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .1009
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DI087YL11"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      155 a    360 c    326 g    157 t     11 others
ORIGIN


  Query Match             77.9%;  Score 861.2;  DB 10;  Length 1009;
  Best Local Similarity   98.2%;  Pred. No. 2.7e-143;
  Matches  874;  Conservative    5;  Mismatches   10;  Indels    1;  Gaps    1;

Qy      1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    120 CGCTGCTCCTGCCGGGTGATGGAAAACCCCAGCCCGGCCGCCGCCCTGGGCAAGGCCCTC 179

Qy     61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 TGCGCTCTCCTCCTGGCCACTCTCGGCGCCGCCGGCCAGCCTCTTGGGGGAGAGTCCATC 239

Qy    121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
          |||:|||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    240 TGTWCCGCCAGAGCCCCGGCCAAATACAGCATCACCTTCACGGGCAAGTGGAGCCAGACG 299

Qy    181 gccttccccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggg 240
          |||||||||||||||||||||||| |||||||||||| ||||||||||||||||||||||
Db    300 GCCTTCCCCAAGCAGTACCCCCTGTCCCGCCCCCCTGCGCAGTGGTCTTCGCTGCTGGGG 359

Qy    241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db    360 GCCGCGCATAGCTCCGACTACAGCATGTGGAGGAAGAACCAGTACGTCAGTMACGGGCTG 419

Qy    301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    420 CGCGACTTTGCGGAGCGCGGCGAGGCCTGGGCGCTGATGAAGGAGATCGAGGCGGCGGGG 479

Qy    361 gaggcgctgcagagcgtgcacgc-ggtgttttcggcgcccgccgtccccagcggcaccgg 419
          |||||||||||||||||||||||| | ||||||||||||||||||||||||||||||||
Db    480 GAGGCGCTGCAGAGCGTGCACGCNGNTGTTTTCGGCGCCCGCCGTCCCCAGCGGCACCGG 539

Qy    420 gcagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcg 479
          |||||||||||||||   | ||||||||||||||||||||||||||||||||||||||
Db    540 GCAGACGTCGGCGGACTGGNAGGTGCAGCGCAGGCACTCGCTGGTCTCGTTTGTGGTGCG 599

Qy    480 catcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacgggga 539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    600 CATCGTGCCCAGCCCCGACTGGTTCGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGGA 659

Qy    540 ccgttggcgggaacaggcggcgctggacctgtaccccctacgacgccgggacggacagcgg 599
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    660 CCGTTGGCGGGAACAGGCGGCGCTGGACCTGTACCCCTACGACGCCGGGACGGACAGCGG 719

Qy    600 cttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataac 659
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    720 CTTCACCTTCTCCTCCCCCAACTTCGCCACCATCCCGCAGGACACGGTGACCGAGATAAC 779

Qy    660 gtcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcc 719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    780 GTCCTCCTCTCCCAGCCACCCGGCCAACTCCTTCTACTACCCGCGGCTGAAGGCCCTGCC 839

Qy    720 tcccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctcc 779
          |||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db    840 TCCCATCGCCAGGGTGACACTGGTGCGGCTGCGACAGARCCCCAGGGCCTTCATCCCTCC 899

Qy    780 cgccccagtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaac 839
          ||||||||| ||||||||||||||||||||||||||||||||||||:||||||||||||
Db    900 CGCCCCAGTTCTGCCCAGCAGGGACAATGAGATTGTAGACAGCGCCTCARTTCCAGAAAC 959

Qy    840 gccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcggag 889
          ||||||||||||||| |||||||||||||||||||||||||||||||:|
```

```
Db    960 GCCGCTGGACTGCGAAGTCTCCCTGTGGTCGTCCTGGGGACTGTGCGGRG 1009


RESULT   4
AL551401
LOCUS       AL551401      904 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL551401 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI064YC22 5
            prime, mRNA sequence.
ACCESSION   AL551401
VERSION     AL551401.1  GI:12889312
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 904)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .904
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DI064YC22"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      145 a    319 c    289 g    148 t      3 others
ORIGIN


  Query Match           75.8%;  Score 837.2;  DB 10;  Length 904;
  Best Local Similarity   98.5%;  Pred. No. 4.8e-139;
  Matches  894;  Conservative    3;  Mismatches    6;  Indels    5;  Gaps    5;

      7 tcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctctgcgct 66
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      1 TCCTGCCGGGTGATGGAAAACCCCAGCCCGGCCGCCGCCCTGGGCAAGGCCCTCTGCGCT 60

Qy    67 ctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatctgttcc 126
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CTCCTCCTGGCCACTCTCGGCGCCGCCGGCCAGCCTCTTGGGGGAGAGTCCATCTGTTCC 120

Qy   127 gccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacggccttc 186
        ||||||:||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GCCAGARCCCCGGCCAAATACAGCATCACCTTCACGGGCAAGTGGAGCCAGACGGCCTTC 180

Qy   187 cccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggggccgcg 246
        |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   181 CCCAAGCAGTACCCCCTGTTCCGCCCCCCTGCGCAGTGGTCTTCGCTGCTGGGGGCCGCG 240

Qy   247 catagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctgcgcgac 306
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CATAGCTCCGACTACAGCATGTGGAGGAAGAACCAGTACGTCAGTAACGGGCTGCGCGAC 300

Qy   307 tttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggggaggcg 366
        |||||||| :|||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TTTGCGGA-SGCGGCGAGGCCTGGGCGCTGATGAAGGAGATCGAGGCGGCGGGGGAGGCG 359

Qy   367 ctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccgggcagacg 426
        |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   360 CTGCAGAGCGTGCACGCGTTGTTTTCGGCGCCCGCCGTCCCCAGCGGCACCGGGCAGACG 419

Qy   427 tcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgcatcgtg 486
        |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   420 TCGGCGGA-CTGGAGGTGCAGCGCAGGCACTCGCTGGTCTCGTTTGTGGTGCGCATCGTG 478

Qy   487 cccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggaccgttgg 546
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   479 CCCAGCCCCGACTGGTTCGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGGACCGTTGG 538

Qy   547 cgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggcttcacc 606
        |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||| |
Db   539 CGGGAACAGGCGGCGCTGGACCTGTACCCCTACGACGCCGGGAC-GACAGCGGCTTCATC 597

Qy   607 ttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacgtcctcc 666
        |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   598 TTCTCCTCCCCCAACTTCGCCACCATCCCGCAGAACACGGTGACCGAGATAACGTCCTCC 657

Qy   667 tctcccagccacccggccaactccttctactacccgcggctgaaggccctgcctcccatc 726
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   658 TCTCCCAGCCACCCGGCCAACTCCTTCTACTACCCGCGGCTGAAGGCCCTGCCTCCCATC 717

Qy   727 gccagggtgacact-ggtgcggctgcgacagagccccagggccttcatccctcccgcccc 785
        ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   718 GCCAGGGTGACACTGGGTGCGGCTGCGACAGAGCCCCAGGGCCTTCATCCCTCCCGCCCC 777

Qy   786 agtcctgcccagcagggacaatgagattgtagacagcgcctcagttccagaaacgccgct 845
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   778 AGTCCTGCCCAGCAGGGACAATGAGATTGTAGACAGCGCCTCAGTTCCAGAAACGCCGCT 837

Qy   846 ggactgcgaggtctccctgtggtcgtcctggggactgtgcggaggccactgtgggaggct 905
        |||||||||||||||||||||||||||||||||||||||||||| |:|||||||||||||
Db   838 GGACTGCGAGGTCTCCCTGTGGTCGTCCTGGGGACTGTGCGGA-GSCACTGTGGGAGGCT 896

Qy   906 cgggacca 913
        |||||| |
Db   897 CGGGACAA 904


RESULT   5
AL551786
LOCUS       AL551786      979 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL551786 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI061YD20 5
            prime, mRNA sequence.
ACCESSION   AL551786
VERSION     AL551786.1  GI:12890071
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 979)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .979
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DI061YD20"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
```

```
                    a division of Invitrogen 9800 Medical Center Drive
                    Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                    Email : fliang@lifetech.com URL :
                    http://fulllength.invitrogen.com"
BASE COUNT      166 a    334 c    307 g    156 t     16 others
ORIGIN

  Query Match             74.8%;  Score 826.8;  DB 10;  Length 979;
  Best Local Similarity   95.5%;  Pred. No. 3.4e-137;
  Matches  941;  Conservative   11;  Mismatches   23;  Indels   10;  Gaps   10;

Qy      34 ccggccgccgccctgggcaaggccctctgcgctctcctcctggccactctcggcgccgcc 93
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CCGGCCGCCGCCCTGGGCAAGGCCCTCTGCGCTCTCCTCCTGGCCACTCTCGGCGCCGCC 60

Qy      94 ggccagcctcttgggggagagtccatctgttccgccagagccctggccaaatacagcatc 153
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db      61 GGCCAGCCTCTTGGGGGAGAGTCCATCTGTTCCGCCAGAGCCCCGGCCAAATACAGCATC 120

Qy     154 accttcacgggcaagtggagccagacggccttccccaagcagtacccccctgttccgcccc 213
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ACCTTCACGGGCAAGTGGAGCCAGACGGCCTTCCCCAAGCAGTACCCCCTGTTCCGCCCC 180

Qy     214 cctgcccagtggtcttcgctgctgggggccgcgcatagctccgactacagcatgtggagg 273
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CCTGCGCAGTGGTCTTCGCTGCTGGGGGCCGCGCATAGCTCCGACTACAGCATGTGGAGG 240

Qy     274 aagaaccagtacgtcagtaacgggctgcgcgactttgcggagcgcggcgaggcctgggcg 333
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     241 AAGAACCAGTACGTCAGTAACGGGCTGCGCGACTTTGCGGAG-GCGGCGAGGCCTGGGCG 299

Qy     334 ctgatgaaggagatcgaggcggcgggggaggcgctgcagagcgtgcacgcggtgttttcg 393
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     300 CTGATGAAGGAGATCGAGGCGGCGGGGGAGGCGCTGCAGAGCGTGCACGAGGTGTTTTCG 359

Qy     394 gcgcccgccgtccccagcggcaccgggcagacgtcggcggagctggaggtgcagcgcagg 453
           |||||||||||||||||||||||||||||||||||||||||||||||| :||||||||||||
Db     360 GCGCCCGCCGTCCCCAGCGGCACCGGGCAGACGTCGGCGGA-STGGAGGTGCAGCGCAGG 418

Qy     454 cactcgctggtctcgtttgtg-gtgcgcatcgtgcccagccccgactggttcgtgggcgt 512
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     419 CACTCGCTGGTCTCGTTTGTGCGTGCGCATCGTGCCCAGCCCCGACTGGTTCGTGGGCGT 478

Qy     513 ggacagcctggacctgtgcgacgggaccgttggcgggaacaggcggcgctggacctgta 572
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     479 GGACAGCCTGGACCTGTGCGACGGGGACCGTTGGCGGGAACAGGCGGCGCTGGACCTGTA 538

Qy     573 cccctacgacgccgggacggacagcggcttcaccttctcctcccccaacttcgccaccat 632
           |||||||| |||||| |||||||||||||| |||||| || | || |||
Db     539 CCCCTACGAAGCCGGGANGGACAGCGGCTTCAACTTCTCCT-CCCCAANTTTGNCANCAT 597

Qy     633 cccgcaggacacggtgaccgagataacgtcctcctctcccagccacccggccaactcctt 692
           |||| | || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     598 CCCGNA-GAAACGGTGACCGAGATAACGTCCTCCTCTCCCAGCCACCCGGCCAACTCCTT 656

Qy     693 ctactacccgcggctgaaggccctgcctcccatcgccagggtgacactggtgcggctgcg 752
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     657 CTACTACCCGCGGCTGAAGGCCCTGCCTCCCATCGCCAGGGTGACACTGCTGCGGCTGCG 716

Qy     753 acagagccccagggccttcatccctcccgccccagtcctgcccagcagggacaatgagat 812
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     717 ACAGAGCCCCAGGGCCTTCATCCCTCCCGCCCCAGTCCTGCCCAGCAGGGACAATGAGAT 776

Qy     813 tg-tagacagcgcctcagttccagaaacgccgctggactgcgaggtctccctgtgg-tcg 870
           || |||||||||||||||||||||||||||||:||||||||||||||||:| |||
Db     777 TGCTAGACAGCGCCTCAGTTCCAGAAACGCCGCTGSACTGCGAGGTCTCCCTGTKGCTCG 836

Qy     871 tcctggggactgtgcggaggccactg-tgggaggctcgggaccaagagcaggactcgcta 929
           ||||||||:|||||||||||| |||| |:| |||||||::|||||||||: |||:|||
Db     837 TCCTGGGGMCTGTGCGGAGGCAACTGCTSGCAGGCTCGSSACCAAGAGCAGSCCTCSCTA 896

Qy     930 cgtccgggtccagcccgccaacaacgggagcccctgccccgagctcgaagaagaggctga 989
           ||:|||||  |||| ||| || || ||||| ||||||||||||||||||||||||||||
Db     897 CGYCCGGGCTCAGCGCGCGAAAAAAGGGAG-CCCTGCCCCGAGCTCGAAGAAGAGGCTGA 955

Qy     990 gtgcgtccctgataactgcgtctaa 1014
           |||||| |||:|||| |||||||||
Db     956 GTGCGT-CCTKATAAATGCGTCTAA 979


RESULT   6
AL554806
LOCUS       AL554806      801 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL554806 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI086YG11 5
            prime, mRNA sequence.
ACCESSION   AL554806
VERSION     AL554806.1  GI:12895943
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 801)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .801
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DI086YG11"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      121 a    299 c    254 g    127 t
ORIGIN

  Query Match             69.0%;  Score 762.6;  DB 10;  Length 801;
  Best Local Similarity   99.2%;  Pred. No. 7.8e-126;
  Matches  787;  Conservative    0;  Mismatches    4;  Indels    2;  Gaps    2;

Qy       1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 CGCTGCTCCTGCCGGGTGATGGAAAACCCCAGCCCGGCCGCCGCCCTGGGCAAGGCCCTC 70

Qy      61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71 TGCGCTCTCCTCCTGGCCACTCTCGGCGCCGCCGGCCAGCCTCTTGGGGGAGAGTCCATC 130

Qy     121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     131 TGTTCCGCCAGAGCCCCGGCCAAATACAGCATCACCTTCACGGGCAAGTGGAGCCAGACG 190

Qy     181 gccttccccaagcagtacccccctgttccgccccccctgcccagtggtcttcgctgctgggg 240
           ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     191 GCCTTCCCCAAGCAGTACCCCCTGTTCCGCCCCCCCTGCGCAGTGGTCTTCGCTGCTGGGG 250

Qy     241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 GCCGCGCATAGCTCCGACTACAGCATGTGGAGGAAGAACCAGTACGTCAGTAACGGGCTG 310
```

```
Qy     301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     311 CGCGACTTTGCGGA-CGCGGCGAGGCCTGGGCGCTGATGAAGGAGATCGAGGCGGCGGGG 369

Qy     361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccgggg 420
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     370 GAGGCGCTGCAGAGCGTGCACGAGGTGTTTTCGGCGCCCGCCGTCCCCAGCGGCACCGGGG 429

Qy     421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     430 CAGACGTCGGCGGA-CTGGAGGTGCAGCGCAGGCACTCGCTGGTCTCGTTTGTGGTGCGC 488

Qy     481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     489 ATCGTGCCCAGCCCCGACTGGTTCGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGGAC 548

Qy     541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     549 CGTTGGCGGGAACAGGCGGCGCTGGACCTGTACCCCTACGACGCCGGGACGGACAGCGGC 608

Qy     601 ttcaccttctcctccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     609 TTCACCTTCTCCTCCCCAACTTCGCCACCATCCCGCAGGACACGGTGACCGAGATAACG 668

Qy     661 tcctcctctcccagccacccggccaactccttctactacccgcggctgaaggccctgcct 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     669 TCCTCCTCTCCCAGCCACCCGGCCAACTCCTTCTACTACCCGCGGCTGAAGGCCCTGCCT 728

Qy     721 cccatcgccagggtgacactggtgcggctgcgacagagccccagggccttcatccctccc 780
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     729 CCCATCGCCAGGGTGACACTGCTGCGGCTGCGACAGAGCCCCAGGGCCTTCATCCCTCCC 788

Qy     781 gccccagtcctgc 793
           |||||||||||||
Db     789 GCCCCAGTCCTGC 801
```

```
RESULT   7
AL543187
LOCUS       AL543187      825 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL543187 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI002YM16 5
            prime, mRNA sequence.
ACCESSION   AL543187
VERSION     AL543187.1  GI:12875665
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 825)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .825
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DI002YM16"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
```

```
                     http://fulllength.invitrogen.com"
BASE COUNT      131 a    292 c    265 g    135 t      2 others
ORIGIN
```

```
  Query Match           63.5%;  Score 701.6;  DB 10;  Length 825;
  Best Local Similarity  99.6%;  Pred. No. 5.1e-115;
  Matches  713;  Conservative    1;  Mismatches    1;  Indels    1;  Gaps    1;

Qy     390 ttcggcgcccgccgtccccagcggcaccgggcagacgtcggcggagctggaggtgcagcg 449
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TTCGGCGCCCGCCGTCCCCAGCGGCACCGGGCAGACGTCGGCGGA-CTGGAGGTGCAGCG 59

Qy     450 caggcactcgctggtctcgtttgtggtgcgcatcgtgcccagccccgactggttcgtggg 509
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60 CAGGCACTCGCTGGTCTCGTTTGTGGTGCGCATCGTGCCCAGCCCCGACTGGTTCGTGGG 119

Qy     510 cgtggacagcctggacctgtgcgacggggaccgttggcgggaacaggcggcgctggacct 569
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 CGTGGACAGCCTGGACCTGTGCGACGGGGACCGTTGGCGGGAACAGGCGGCGCTGGACCT 179

Qy     570 gtacccctacgacgccgggacggacagcggcttcaccttctcctccccaacttcgccac 629
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180 GTACCCCTACGACGCCGGGACGGACAGCGGCTTCACCTTCTCCTCCCCAACTTCGCCAC 239

Qy     630 catcccgcaggacacggtgaccgagataacgtcctcctctcccagccacccggccaactc 689
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     240 CATCCCGCAGGACACGGTGACCGAGATAACGTCCTCCTCTCCCAGCCACCCGGCCAACTC 299

Qy     690 cttctactacccgcggctgaaggccctgcctcccatcgccagggtgacactggtgcggct 749
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 CTTCTACTACCCGCGGCTGAAGGCCCTGCCTCCCATCGCCAGGGTGACACTGGTGCGGCT 359

Qy     750 gcgacagagccccagggccttcatccctcccgccccagtcctgcccagcagggacaatga 809
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 GCGACAGAGCCCCAGGGCCTTCATCCCTCCCGCCCCAGTCCTGCCCAGCAGGGACAATGA 419

Qy     810 gattgtagacagcgcctcagttccagaaacgccgctggactgcgaggtctccctgtggtc 869
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 GATTGTAGACAGCGCCTCAGTTCCAGAAACGCCGCTGGACTGCGAGGTCTCCCTGTGGTC 479

Qy     870 gtcctggggactgtgcggaggccactgtgggaggctcgggaccaagagcaggactcgcta 929
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 GTCCTGGGGACTGTGCGGAGGCCACTGTGGGAGGCTCGGGACCAAGAGCAGGACTCGCTA 539

Qy     930 cgtccgggtccagcccgccaacaacgggagcccctgcccgagctcgaagaagaggctga 989
           |||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db     540 CGTCCGGGTCCAGCCCGCCAACAACGGGAGCHCCTGCCCCGAGCTCGAAGAAGAGGCTGA 599

Qy     990 gtgcgtccctgataactgcgtctaagaccagagccccgcagcccctggggccccccggag 1049
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 GTGCGTCCCTGATAACTGCGTCTAAGACCAGAGCCCCGCAGCCCCTGGGGCCCCCCGGAG 659

Qy    1050 ccatggggtgtcgggggctcctgtgcaggctcatgctgcaggcggccgagggcaca 1105
           |||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     660 CCATGGGGTGTCGGGGGCTCCTGTGCAGGCTCATACTGCAGGCGGCCGAGGGCACA 715
```

```
RESULT   8
AL550317
LOCUS       AL550317      971 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL550317 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI039YK16 5
            prime, mRNA sequence.
ACCESSION   AL550317
VERSION     AL550317.1  GI:12887174
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 971)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
```

```
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .971
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DI039YK16"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      147 a    354 c    315 g    152 t      3 others
ORIGIN
```

```
  Query Match             62.0%;  Score 685.2;  DB 10;  Length 971;
  Best Local Similarity   98.9%;  Pred. No. 4.2e-112;
  Matches  731;  Conservative    0;  Mismatches    4;  Indels    4;  Gaps    4;

Qy       1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     235 CGCTGCTCCTGCCGGGTGATGGAAAACCCCAGCCCGGCCGCCGCCCTGGGCAAGGCCCTC 294

Qy      61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     295 TGCGCTCTCCTCCTGGCCACTCTCGGCGCCGCCGGCCAGCCTCTTGGGGGAGAGTCCATC 354

Qy     121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     355 TGTTCCGCCAGAGCCCCGGCCAAATACAGCATCACCTTCACGGGCAAGTGGAGCCAGACG 414

Qy     181 gccttccccaagcagtacccccctgttccgccccccctgcccagtggtcttcgctgctgggg 240
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     415 GCCTTCCCCAAGCAGTACCCCCCTGTTCCGCCCCCCTGCGCAGTGGTCTTCGCTGCTGGGG 474

Qy     241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     475 GCCGCGCATAGCTCCGACTACAGCATGTGGAGGAAGAACCAGTACGTCAGTAACGGGCTG 534

Qy     301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     535 CGCGACTTTGCGGAGCGCGGCGAGGCCTGGGCGCTGATGAAGGAGATCGAGGCGGCGGGG 594

Qy     361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
           |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db     595 GAGGCGCTGCAGAGCGTGCACGCGNTGTTTTCGGCGCCCGCCGTCCCCAGCGGCACCGGG 654

Qy     421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     655 CAGACGTCGGCGGA-CTGGAGGTGCAGCGCAGGCACTCGCTGGTCTCGTTTGTGGTGCGC 713

Qy     481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     714 ATCGTGCCCAGCCCCGACTGGTTCGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGGA- 772

Qy     541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     773 CGTTGGCGGGAACAGGCGGCGCTGGACCTGTACCCCTACGACGCCGGGACGGACAGCGGC 832

Qy     601 ttcaccttctcctccccccaacttcgccaccatcccgcaggac-acggtgaccgagataac 659
           |||||||||||||||||||||||||||||||||||| || ||||||||||||||||||
Db     833 TTCACCTTCTCCTCCCCCAACTTCGCCACCATCCCGCAGAACAACGGTGACCGAGATAAC 892
```

```
Qy     660 gtcctcctctcccagccacccggccaact-ccttctactacccgcggctgaaggccctgc 718
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     893 GTCCTCCTCTCCCAGCCACCCGGCCAACTCCCTTCTACTACCCGCGGCTGAAGGCCCTGC 952

Qy     719 ctcccatcgccagggtgac 737
           ||||||||||||||||||||
Db     953 CTCCCATCGCCAGGGTGAC 971


RESULT   9
AL547912
LOCUS       AL547912      931 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL547912 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI033YF20 5
            prime, mRNA sequence.
ACCESSION   AL547912
VERSION     AL547912.1  GI:12882420
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 931)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .931
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DI033YF20"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      142 a    336 c    295 g    153 t      5 others
ORIGIN
```

```
  Query Match             59.1%;  Score 653.6;  DB 10;  Length 931;
  Best Local Similarity   97.4%;  Pred. No. 1.6e-106;
  Matches  711;  Conservative    5;  Mismatches    9;  Indels    5;  Gaps    5;

Qy      12 ccgggtgatggaaaacccca-gcccggccgccgccctgggcaaggccctctgcgctctcc 70
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     194 CCGGGTGATGGAAAACCCCACGCCCGGCCGCCGCCCTGGGCAAGGCCCTCTGCGCTCTCC 253

Qy      71 tcctggccactctcggcgccgccggccagcctcttgggggagagtccatctgttccgcca 130
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     254 TCCTGGCCACTCTCGGCGCCGCCGGCCAGCCTCTTGGGGGAGAGTCCATCTGTTCCGCCA 313

Qy     131 gagccctggccaaatacagcatcaccttcacgggcaagtggagccagacggccttcccca 190
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     314 GAGCCCCGGCCAAATACAGCATCACCTTCACGGGCAAGTGGAGCCAGACGGCCTTCCCCA 373

Qy     191 agcagtacccccctgttccgccccccctgcccagtggtcttcgctgctgggggccgcgcata 250
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     374 AGCAGTACCCCCTGTTCCGCCCCCCTGCGCAGTGGTCTTCGCTGCTGGGGGCCGCGCATA 433

Qy     251 gctccgactacagcatgtggaggaagaacc-agtacgtcagtaacgggctgcgcgacttt 309
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
```

```
Db    434 GCTCCGACTACAGCATGTGGAGGAAGAACCTAGTACGTCAGTAACGGGCTGCGCGACTTT 493

Qy    310 gcggagcgcggcgaggcctgggcgctgatgaag-gagatcgaggcggcgggggaggcgct 368
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    494 GCGGAGCGCGGCGAGGCCTGGGCGCTGATGAAGAGAGATCGAGGCGGCGGGGGAGGCGCT 553

Qy    369 gcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccgggcagacgtc 428
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    554 GCAGAGCGTGCACGCGATGTTTTCGGCGCCCGCCGTCCCCAGCGGCACCGGGCAGACGTC 613

Qy    429 ggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgcatcgtgcc 488
          ||||||   | ||||||||||||||||||||||||||||||||||||||||||||||||
Db    614 GGCGGACTGGAAGGTGCAGCGCAGGCACTCGCTGGTCTCGTTTGTGGTGCGCATCGTGCC 673

Qy    489 cagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggaccgttggcg 548
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    674 CAGCCCCGACTGGTTCGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGGACCGTTGGCG 733

Qy    549 ggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggcttcaccttt 608
          ||||:|:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    734 GGAAMARGCGGCGCTGGACCTGTACCCCTACGACGCCGGGACGGACAGCGGCTTCACCTT 793

Qy    609 ctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacgtcctcctc 668
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    794 CTCCTCCCCCAACTTCGCCACCATCCCGCAGAACACGGTGACCGAGATAACGTCCTCCTC 853

Qy    669 tcccagccacccggccaactccttctactacccgcggctgaaggccctgcctcccatcgc 728
          |||||||:| ||||:|||||||||||||||||||||||| ||||||||||||||||||||
Db    854 TCCCAGCMA-CCGGSCAACTCCTTCTACTACCCGCGGCTGAAGG-CCTGCCTCCCATCGC 911

Qy    729 cagggtgaca 738
          :||||||| |
Db    912 MAGGGTGAAA 921
```

RESULT 10
AL532820
LOCUS AL532820 935 bp mRNA EST 13-FEB-2001
DEFINITION AL532820 LTI_FL015_Brn1 Homo sapiens cDNA clone CS0DN002YN20 5 prime, mRNA sequence.
ACCESSION AL532820
VERSION AL532820.1 GI:12796313
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 935)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES Location/Qualifiers
source 1. .935
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DN002YN20"
/clone_lib="LTI_FL015_Brn1"
/sex="male"
/tissue_type="Adult brain"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 138 a 328 c 312 g 145 t 12 others
ORIGIN

Query Match 56.2%; Score 621.4; DB 10; Length 935;
Best Local Similarity 97.2%; Pred. No. 8.4e-101;
Matches 652; Conservative 10; Mismatches 6; Indels 3; Gaps 3;

```
Qy      1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 CGCTGCTCCTGCCGGGTGATGGAAAACCCCAGCCCGGCCGCCGCCCTGGGCAAGGCCCTC 320

Qy     61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 TGCGCTCTCCTCCTGGCCACTCTCGGCGCCGCCGGCCAGCCTCTTGGGGGAGAGTCCATC 380

Qy    121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||:|
Db    381 TGTTCCGCCAGAGCCCCGGCCAAATACAGCATCACCTTCACGGGCAAGTGGAGCCAGAMG 440

Qy    181 gccttccccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggg 240
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    441 GCCTTCCCCAAGCAGTACCCCCTGTTCCGCCCCCCTGCGCAGTGGTCTTCGCTGCTGGGG 500

Qy    241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    501 GCCGCGCATAGCTCCGACTACAGCATGTGGAGGAAGAACCAGTACGTCAGTTACGGGCTG 560

Qy    301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    561 CGCGACTTTGCGGA-CGCGGCGAGGCCTGGGCGCTGATGAAGGAGATCGAGGCGGCGGGG 619

Qy    361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
          |||||||||||||||||||| |||||||||||||||| ||||||||||||||||||||||
Db    620 GAGGCGCTGCAGAGCGTGCACGAGGTGTTTTCGGCGCCCG-CGTCCCCAGCGGCACCGGG 678

Qy    421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
          ||||||||||||||:  ||||||||||||||||||||||||||||||||||||||||||
Db    679 CAGACGTCGGCGGASTGGGAGGTGCAGCGCAGGCACTCGCTGGTCTCGTTTGTGGTGCGC 738

Qy    481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    739 ATCGTGCCCAGCCCCGACTGGTTCGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGGAC 798

Qy    541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
          ||||||||||||||||||:|||||||||||:|||||:||:|||||||||||:||:||||
Db    799 CGTTGGCGGGAACAGGCGGSGCTGGACCTGTMCCCCTMCGMCGCCGGGACGGMCASCGGC 858

Qy    601 ttcaccttctcctccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
          ||||||||:|||| ||||||||||||||||||||||||||||||||||||||:||||
Db    859 TTCACCTTYTCCT-CCCCAACTTCGCCACCATCCCGCAGGACACGGTGACCGAGAWAACG 917

Qy    661 tcctcctctcc 671
          |||||||||||
Db    918 TCCTCCTCTCC 928
```

RESULT 11
AL564981/c
LOCUS AL564981 989 bp mRNA EST 16-FEB-2001
DEFINITION AL564981 LTI_FL015_Brn1 Homo sapiens cDNA clone CS0DN002YN20 3 prime, mRNA sequence.
ACCESSION AL564981
VERSION AL564981.1 GI:12915901
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 989)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)

```
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .989
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DN002YN20"
                     /clone_lib="LTI_FL015_Brn1"
                     /sex="male"
                     /tissue_type="Adult brain"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was constructed by Life Technologies.
                     Contact : Feng Liang Life Technologies, a division of
                     Invitrogen 9800 Medical Center Drive Rockville, Maryland
                     20850, USA Fax : (1) 301 610 8371 Email :
                     fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      196 a    278 c    338 g    154 t     23 others
ORIGIN
```

```
  Query Match             55.5%;  Score 613.8;  DB 10;  Length 989;
  Best Local Similarity   98.3%;  Pred. No. 1.9e-99;
  Matches  626;  Conservative    4;  Mismatches    6;  Indels    1;  Gaps    1;

Qy    470 ttgtggtgcgcatcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgt 529
          || | ||||||:||||||||||||||| |||||||||||||||| ||||||||||||||||
Db    986 TTTTTGTGCGCWTCGTGCCCAGCCCCAACTGGTTCGTGGGCGTGAACAGCCTGGACCTGT 927

Qy    530 gcgacggggaccgttggcgggaacaggcggcgctggacctgtacccctacgacgccggga 589
          ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    926 GCRACGGGGACCGTTGGCGGGAACAGGCGGCGCTGGACCTGTACCCCTACGACGCCGGGA 867

Qy    590 cggacagcggcttcaccttctcct-cccccaacttcgccaccatcccgcaggacacggtg 648
          ||:|||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    866 CGRACAGCGGCTTCACCTTCTCCTCCCCCCAACTTCGCCACCATCCCGCAGGACACGGTG 807

Qy    649 accgagataacgtcctcctctcccagccacccggccaactccttctactacccgcggctg 708
          ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||
Db    806 ACCGAGATAACGTCCTCCTCTCCCAGCCACCCGGCCAACTCCTTCTACWACCCGCGGCTG 747

Qy    709 aaggccctgcctcccatcgccagggtgacactggtgcggctgcgacagagccccagggcc 768
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    746 AAGGCCCTGCCTCCCATCGCCAGGGTGACACTGCTGCGGCTGCGACAGAGCCCCAGGGCC 687

Qy    769 ttcatccctcccgccccagtcctgcccagcagggacaatgagattgtagacagcgcctca 828
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    686 TTCATCCCTCCCGCCCCAGTCCTGCCCAGCAGGGACAATGAGATTGTAGACAGCGCCTCA 627

Qy    829 gttccagaaacgccgctggactgcgaggtctccctgtggtcgtcctggggactgtgcgga 888
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    626 GTTCCAGAAACGCCGCTGGACTGCGAGGTCTCCCTGTGGTCGTCCTGGGGACTGTGCGGA 567

Qy    889 ggccactgtgggaggctcgggaccaagagcaggactcgctacgtccgggtccagcccgcc 948
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    566 GGCCACTGTGGGAGGCTCGGGACCAAGAGCAGTACTCGCTACGTCCGGGTCCAGCCCGCC 507

Qy    949 aacaacgggagcccctgccccgagctcgaagaagaggctgagtgcgtccctgataactgc 1008
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    506 AACAACGGGAGCCCCTGCCCCGAGCTCGAAGAAGAGGCTGAGTGCGTCCCTGATAACTGC 447

Qy   1009 gtctaagaccagagccccgcagcccctggggcccccccggagccatggggtgtcgggggct 1068
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    446 GTCTAAGACCAGAGCCCCGCAGCCCCTGGGGCCCCCCCGGAGCCATGGGGTGTCGGGGGCT 387

Qy   1069 cctgtgcaggctcatgctgcaggcggccgagggcaca 1105
          |||||||||||||||||||||||||||||||||||||
Db    386 CCTGTGCAGGCTCATGCTGCAGGCGGCCGAGGGCACA 350
```

```
RESULT  12
AL551123
LOCUS       AL551123      838 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL551123 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI066YN19 5
            prime, mRNA sequence.
ACCESSION   AL551123
VERSION     AL551123.1  GI:12888764
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 838)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .838
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DI066YN19"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      122 a    296 c    281 g    134 t      5 others
ORIGIN
```

```
  Query Match             54.4%;  Score 600.6;  DB 10;  Length 838;
  Best Local Similarity   98.4%;  Pred. No. 4e-97;
  Matches  614;  Conservative    2;  Mismatches    7;  Indels    1;  Gaps    1;

Qy      1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgcccctgggcaaggccctc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    216 CGCTGCTCCTGCCGGGTGATGGAAAACCCCAGCCCGGCCGCCGCCCTGGGCAAGGCCCTC 275

Qy     61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    276 TGCGCTCTCCTCCTGGCCACTCTCGGCGCCGCCGGCCAGCCTCTTGGGGGAGAATCCATC 335

Qy    121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    336 TGTTCCGCCAGAGCCCCGGCCAAATACAGCATCACCTTCACGGGCAAGTGGAGCCAGACG 395

Qy    181 gccttccccaagcagtacccccctgttccgccccccctgcccagtggtcttcgctgctgggg 240
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    396 GCCTTCCCCAAGCAGTACCCCCTGTTCCGCCCCCCTGCGCAGTGGTCTTCGCTGCTGGGG 455

Qy    241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    456 GCCGCGCATAGCTCCGACTACAGCATGTGGAGGAAGAACCAGTACGTCAGTAACGGGCTG 515

Qy    301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    516 CGCGACTTTGCGGAG-GCGGCGAGGCCTGGGCGCTGATGAAGGAGATCGAGGCGGCGGGG 574

Qy    361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
```

```
Db    575 GAGGCGCTGCAGAGCGTGCACGCGNTGTTTTCGGCGCCCGCCGTCCCCAGCGGCACCGGG 634

Qy    421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
          |||||||||||||||:   ||||||||||||||||||||||||||||||||||||||||||
Db    635 CAGACGTCGGCGGASTGTGAGGTGCAGCGCAGGCACTCGCTGGTCTCGTTTGTGGTGCGC 694

Qy    481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    695 ATCGTGCCCAGCCCCGACTGGTTCGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGGAC 754

Qy    541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    755 CGTTGGCGGGAACAGGCGGCGCTGGACCTGTACCCCTACGACGCCGGGACGGACAGCGGC 814

Qy    601 ttcaccttctcctccccaacttc 624
          |||||||||||||:||||||||||
Db    815 TTCACCTTCTCCTSCCCCAACTTC 838


RESULT 13
BE407510
LOCUS       BE407510      682 bp    mRNA            EST       21-JUL-2000
DEFINITION  601300346F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3630467 5',
            mRNA sequence.
ACCESSION   BE407510
VERSION     BE407510.1  GI:9343960
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 682)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM318  row: k  column: 12
            High quality sequence stop: 648.
FEATURES             Location/Qualifiers
     source          1. .682
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3630467"
                     /clone_lib="NIH_MGC_21"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      108 a    237 c    230 g    107 t
ORIGIN


  Query Match           54.3%;  Score 600.4;  DB 10;  Length 682;
  Best Local Similarity  96.2%;  Pred. No. 4.2e-97;
  Matches  637;  Conservative    0;  Mismatches   21;  Indels    4;  Gaps    2;

Qy      1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgccctgggcaaggccctc 60
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     23 CGCTGCTCCTGCCGGGTGATGGAAAACCCCAGCCCGGGCGCCGCCCTGGGCAAGGCCCTC 82

Qy     61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     83 TGCGCTCTCCTCCTGGCCACTCTCGGCGCCGCCGGCCAGCCTCTTGGGGGAGAGTCCATC 142

Qy    121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    143 TGTTCCGCCAGAGCCCCGGCCAAATACAGCATCACCTTCACGGGCAAGTGGAGCCAGACG 202

Qy    181 gccttccccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggg 240
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    203 GCCTTCCCCAAGCAGTACCCCCTGTTCCGCCCCCCTGCGCAGTGGTCTTCGCTGCTGGGG 262

Qy    241 gccgcgcatagctccgactacagcatgtggaggaagaaccagtacgtcagtaacgggctg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    263 GCCGCGCATAGCTCCGACTACAGCATGTGGAGGAAGAACCAGTACGTCAGTAACGGGCTG 322

Qy    301 cgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgggg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    323 CGCGACTTTGCGGAGCGCGGCGAGGCCTGGGCGCTGATGAAGGAGATCGAGGCGGCGGGG 382

Qy    361 gaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccggg 420
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    383 GAGGCGCTGCAGAGCGTGCACGAGGTGTTTTCGGCGCCCGCCGTCCCCAGCGGCACCGGG 442

Qy    421 cagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgcgc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    443 CAGACGTCGGCGGAGCTGGAGGTGCAGCGCAGGCACTCGCTGGTCTCGTTTGTGGTGCGC 502

Qy    481 atcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacggggac 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    503 ATCGTGCCCAGCCCCGACTGGTTCGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGGAC 562

Qy    541 cgttggcgggaacaggcggcgctggacctgtacccctacgacgccgggacggacagcggc 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    563 CGTTGGCGGGAACAGGCGGCGCTGGACCTGTACCCCTACGACGCCGGGACGGACAGCGG- 621

Qy    601 ttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgagataacg 660
          |   ||||||||||||||||  |   || ||||||||||   |||   ||    ||||||
Db    622 -TTCACTTCTCCTCCCCCAA--TTCGCAACATCCCGCAGACAACGTGAACGAGAATAACG 678

Qy    661 tc 662
          ||
Db    679 TC 680


RESULT 14
AL574221/c
LOCUS       AL574221     1025 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL574221 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI039YK16 3
            prime, mRNA sequence.
ACCESSION   AL574221
VERSION     AL574221.1  GI:12934220
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1025)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .1025
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DI039YK16"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
```

was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 192 a 306 c 343 g 176 t 8 others
ORIGIN

Query Match 53.7%; Score 593; DB 10; Length 1025;
Best Local Similarity 99.2%; Pred. No. 9.1e-96;
Matches 604; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

```
Qy    497 actggttcgtgggcgtggacagcctggacctgtgcgacggggaccgttggcgggaacagg 556
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1024 ACTGGTTCGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGGACCGTTGGCGGGAACAGG 965

Qy    557 cggcgctggacctgtacccctacgacgccgggacggacagcggcttcaccttctcctccc 616
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    964 CGGCGCTGGACCTGTACCCCTACGACGCCGGGACGGACAGC-GCTTCACCTTCTCCTCCC 906

Qy    617 ccaacttcgccaccatcccgcaggacacggtgaccgagataacgtcctcctctcccagcc 676
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    905 CAAACTTCGCCACCATCCCGCAGGACACGGTGACCGAGATAACGTCCTCCTCTCCCAGCC 846

Qy    677 acccggccaactccttctactacccgcggctgaaggccctgcctcccatcgccagggtga 736
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    845 ACCCGGCCAACTCCTTCTACTACCCGCGGCTGAAGGCCCTGCCTCCCATCGCCAGGGTGA 786

Qy    737 cactggtgcggctgcgacagagccccagggccttcatcctcccgccccagtcctgccca 796
          ||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db    785 CACTGGTGCGGCTGCGACAGAGCCCCAGGGCCTTCATCCSTCCCGCCCCAGTCCTGCCCA 726

Qy    797 gcagggacaatgagattgtagacagcgcctcagttccagaaacgccgctggactgcgagg 856
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    725 GCAGGGACAATGAGATTGTAGACAGCGCCTCAGTTCCAGAAACGCCGCTGGACTGCGAGG 666

Qy    857 tctccctgtggtcgtcctggggactgtgcggaggccactgtgggaggctcgggaccaaga 916
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    665 TCTCCCTGTGGTCGTCCTGGGGACTGTGCGGAGGCCACTGTGGGAGGCTCGGGACCAAGA 606

Qy    917 gcaggactcgctacgtccgggtccagcccgccaacaacgggagcccctgccccgagctcg 976
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    605 GCAGGACTCGCTACGTCCGGGTCCAGCCCGCCAACAACGGGAGCCCCTGCCCCGAGCTCG 546

Qy    977 aagaagaggctgagtgcgtccctgataactgcgtctaagaccagagccccgcagcccctg 1036
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    545 AAGAAGAGGCTGAGTGCGTCCCTGATAACTGCGTCTAAGACCAGAGCCCCGCAGCCCCTG 486

Qy   1037 gggccccccggagccatggggtgtcgggggctcctgtgcaggctcatgctgcaggcggcc 1096
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    485 GGGCCCCCCGGAGCCATGGGGTGTCGGGGGCTCCTGTGCAGGCTCATACTGCAGGCGGCC 426

Qy   1097 gagggcaca 1105
          |||||:|||
Db    425 GAGGGSACA 417
```

RESULT 15
AL542564
LOCUS AL542564 904 bp mRNA EST 16-FEB-2001
DEFINITION AL542564 LTI_FL002_PL1 Homo sapiens cDNA clone CS0DE012YP18 5 prime , mRNA sequence.
ACCESSION AL542564
VERSION AL542564.1 GI:12874732
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 904)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES Location/Qualifiers
source 1. .904
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DE012YP18"
/clone_lib="LTI_FL002_PL1"
/lab_host="DH10B"
/note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville , Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 136 a 316 c 306 g 143 t 3 others
ORIGIN

Query Match 52.4%; Score 579.2; DB 10; Length 904;
Best Local Similarity 97.1%; Pred. No. 2.5e-93;
Matches 639; Conservative 3; Mismatches 11; Indels 5; Gaps 5;

```
Qy      1 cgctgctcctgccgggtgatggaaaaccccagcccggccgccgcccctgggcaaggccctc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    249 CGCTGCTCCTGCCGGGTGATGGAAAACCCCAGCCCGGCCGCCGCCCTGGGCAAGGCCCTC 308

Qy     61 tgcgctctcctcctggccactctcggcgccgccggccagcctcttgggggagagtccatc 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    309 TGCGCTCTCCTCCTGGCCACTCTCGGCGCCGCCGGCCAGCCTCTTGGGGGAGAATCCATC 368

Qy    121 tgttccgccagagccctggccaaatacagcatcaccttcacgggcaagtggagccagacg 180
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    369 TGTTCCGCCAGAGCCCCGGCCAAATACAGCATCACCTTCACGGGCAAGTGGAGCCAGACG 428

Qy    181 gccttccccaagcagtacccccctgttccgcccccctgcccagtggtcttcgctgctgggg 240
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    429 GCCTTCCCCAAGCAGTACCCCCTGTTCCGCCCCCCTGCGCAGTGGTCTTCGCTGCTGGGG 488

Qy    241 gccgcgcata-gctccgacta-cagcatgtggaggaagaaccagtacgtcagtaacgggc 298
          |||||||||| |||||||||| |||||||||||||||||||||||||||||| ||||||
Db    489 GCCGCGCATATGCTCCGACTATCAGCATGTGGAGGAAGAACCAGTACGTCAGTTACGGGC 548

Qy    299 tgcgcgactttgcggagcgcggcgaggcctgggcgctgatgaaggagatcgaggcggcgg 358
          |||||||||||||||| :|||||||||||||||||||||||||||||||||||||||||
Db    549 TGCGCGACTTTGCGGA-SGCGGCGAGGCCTGGGCGCTGATGAAGGAGATCGAGGCGGCGG 607

Qy    359 gggaggcgctgcagagcgtgcacgcggtgttttcggcgcccgccgtccccagcggcaccg 418
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    608 GGGAGGCGCTGCAGAGCGTGCACGCGGTGTTTTCGGCGCCCG-CGTCCCCAGCGGCACCG 666

Qy    419 ggcagacgtcggcggagctggaggtgcagcgcaggcactcgctggtctcgtttgtggtgc 478
          ||||||||||||||||:  ||||||||||||||||||||||||||||||||||||||||
Db    667 GGCAGACGTCGGCGGASTGGGAGGTGCAGCGCAGGCACTCGCTGGTCTCGTTTGTGGTGC 726

Qy    479 gcatcgtgcccagccccgactggttcgtgggcgtggacagcctggacctgtgcgacgggg 538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    727 GCATCGTGCCCAGCCCCGACTGGTTCGTGGGCGTGGACAGCCTGGACCTGTGCGACGGGG 786

Qy    539 accgttggcgggaacaggcggcgctggacctgtacccc-tacgacgccgggacggacagc 597
          ||  | ||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    787 ACGTTGGGCGGGAACAGGCGGCGCTGGACCTGTACCCCTTACGACGCCGGGACGGACAGC 846
```

```
Qy    598 ggcttcaccttctcctcccccaacttcgccaccatcccgcaggacacggtgaccgaga 655
          ||||||||||||||||||||||||||||||||||||||||||||  :||||||||||||
Db    847 GGCTTCACCTTCTCCTCCCCCAACTTCGCCACCATCCCGCAGACMACGGTGACCGAGA 904
```

Search completed: December 10, 2001, 19:12:41
Job time: 20116 sec